

	QY	1 KFLAWWKITYKD 12 :	
RESULT	10	PCT-US97-07950-375 STANDARD; PRT;	111 AA.
ID XX AC XX DD DE EE FF GG HH II JJ KK LL MM NN OO PP QQ RR SS TT UU VV WW YY ZZ	XX	XXXXXX	
Sequence 375,	Application PC/TUS9707950		
Sequence 375,	Appication PC/TUS9707950		
GENERAL INFORMATION:			
APPLICANT:	Black, Michael		
APPLICANT:	Hodgson, John		
APPLICANT:	Knowles, David		
APPLICANT:	Nicholas, Richard		
APPLICANT:	Stodola, Robert		
TITLE OF INVENTION:	Novel Compounds		
NUMBER OF SEQUENCES:	552		
CORRESPONDENCE ADDRESS:			
ADDRESSEE:	SmithKline Beecham Corporation		
STREET:	709 Swedeland Road		
CITY:	King of Prussia		
STATE:	PA		
COUNTRY:	USA		
ZIP:	19406-0939		
COMPUTER READABLE FORM:			
MEDIUM TYPE:	Diskette		
COMPUTER:	IBM Compatible		
OPERATING SYSTEM:	DOS		
SOURCE SOFTWARE:	Fastsq for Windows Version 2.0		
CURRENT APPLICATION DATA:			
APPLICATION NUMBER:	PCT/US97/07950		
FILING DATE:	14-MAY-1997		
CLASSIFICATION:			
PRIOR APPLICATION DATA:			
APPLICATION NUMBER:	60/017670		
FILING DATE:	14-MAY-1996		
ATTORNEY/AGENT INFORMATION:			
NAME:	Gimmi, Edward R		
REGISTRATION NUMBER:	38,891		
REFERENCE/DOCKET NUMBER:	P50475		
TELECOMMUNICATION INFORMATION:			
TELEPHONE:	610-270-4478		
TELEX:	610-270-5090		
INFORMATION FOR SEQ ID NO:	375:		
SEQUENCE CHARACTERISTICS:			
LENGTH:	111 amino acids		
TYPE:	amino acid		
STRANDEDNESS:	single		
TOPOLOGY:	linear		
MOLECULE TYPE:	Protein		
SEQ SEQUENCE	111 AA; 13251 MW; 58279 CN;		
Query Match	59.6%; Score 62; DB 1; Length		
Best Local Similarity	60.0%; Pred. No. 7.42e+01;		
Matches	6; Conservative 2; Mismatches 2;		
Db	66 KYLSA.WITTY 75		
I.I.: I IIII			
QY	1 KFLAVWIKITY 10		
RESULT	11	US-08-858-207A-375 STANDARD; PRT;	111 AA.
ID XX AC XX DD DE EE FF GG HH II JJ KK LL MM NN OO PP QQ RR SS TT UU VV WW YY ZZ	XX	xxxxxx	

CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: GENOME THERAPEUTICS CORPORATION
CC STREET: 100 Beaver Street
CC CITY: Waltham
CC STATE: Massachusetts
CC COUNTRY: USA
CC ZIP: 02354
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: CD-ROM ISO9660
CC COMPUTER:
CC OPERATING SYSTEM:
CC SOFTWARE:
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/09/107,532
CC FILING DATE:
CC PRIOR APPLICATION NUMBER: 60/ 085598
CC FILING DATE: May 14, 1998
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 60/051571
CC FILING DATE: July 2, 1997
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Arieliello, Pamela Deneke
CC REGISTRATION NUMBER: 40,489
CC REFERENCE/DOCKET NUMBER: GTC-012
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (781)893-5007
CC TELEFAX: (781)893-8277
CC INFORMATION FOR SEQ ID NO: 4683:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 606 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC HYPOTHETICAL: YES
CC ORIGINAL SOURCE:
CC ORGANISM: Enterococcus faecium
CC FEATURE:
CC NAME/KEY: misc_feature
CC LOCATION: 1...606
CC
SQ SEQUENCE 606 AA; 70613 MW; 2071988 CN;

Query Match 59.5%; Score 62; DB 15; Length 606;
Best Local Similarity 54.5%; Pred. No. 7.42e-01;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 255 FLTNKIPKIPD 265
||: ||| : |
QY 2 FLAVWKITYKD 12

RESULT 13
ID US-09-134-000-5353 STANDARD; PRT; 80 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 5353, Application US/09134000A
XX
CC Sequence 5353, Application US/09134000A
CC GENERAL INFORMATION:
CC APPLICANT: Lynn Doucette-Stamm et al
CC TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS
CC TITLE OF INVENTION: FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
CC FILE REFERENCE: GTC-005
CC CURRENT APPLICATION NUMBER: US/09/134,000A
CC CURRENT FILING DATE: 1998-08-13
CC NUMBER OF SEQ ID NOS: 6810
CC SEQ ID NO 5353
CC LENGTH: 80
CC TYPE: PRT
CC ORGANISM: Enterococcus faecalis

SQ SEQUENCE 80 AA; 9450 MW; 34784 CN;

Query Match 58.7%; Score 61; DB 16; Length 80;
Best Local Similarity 53.6%; Pred. No. 9.33e-01;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 41 KFLNIRKITYK 51
||| :|||
QY 1 KFLAVWKITYK 11

RESULT 14
ID US-09-253-682-25 STANDARD; PRT; 642 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 25, Application US/09253682
XX
CC Sequence 25, Application US/09253682
CC GENERAL INFORMATION:
CC APPLICANT: Spaete, Richard
CC APPLICANT: Cha, Tai-An
CC TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS
CC NUMBER OF SEQUENCES: 27
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Luann Cserr Attorney at Law
CC STREET: 750 Arimo Avenue
CC CITY: Oakland
CC STATE: CA
CC COUNTRY: USA
CC ZIP: 94610
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/09/253,682
CC FILING DATE:
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/926,922
CC FILING DATE: September 10, 1997
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Cserr, Luann
CC REGISTRATION NUMBER: 31,822
CC REFERENCE/DOCKET NUMBER: AVIR 11A
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 510-834-1448
CC TELEFAX: 510-839-7810
CC INFORMATION FOR SEQ ID NO: 25:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 642 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC IMMEDIATE SOURCE:
CC CLONE: tol.21
CC FEATURE:
CC NAME/KEY: Protein
CC LOCATION: 1..642
CC OTHER INFORMATION: /label= ULI150
SQ SEQUENCE 642 AA; 70848 MW; 2000477 CN;

Query Match 58.7%; Score 61; DB 17; Length 642;
Best Local Similarity 53.6%; Pred. No. 9.33e+01;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 404 LAIWQTYNDT 414
||: ||| :|||
QY 3 LAVWKITYKDT 13

RESULT 15
ID US-08-812-716-21 STANDARD; PRT; 642 AA.
XX
AC xxxxxx
XX
DT
XX
XX
DE Sequence 21, Application US/08812716
XX
XX Sequence 21, Application US/08812716
CC GENERAL INFORMATION:
CC APPLICANT: KEMBLE, George
CC APPLICANT: DUKE, Gregory
CC APPLICANT: SPAETE, Richard
CC TITLE OF INVENTION: ATTENUATION OF CYTOMEGALOVIRUS
CC TITLE OF INVENTION: VIRULENCE
CC NUMBER OF SEQUENCES: 22
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: AVIRON
CC STREET: 297 N. Bernardo Avenue
CC CITY: Mountain View
CC STATE: California
CC COUNTRY: USA
CC ZIP: 94043
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: DOS
CC SOFTWARE: FastSeq for Windows Version 2.0
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/812,716
CC FILING DATE: 06-MAR-1997
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER:
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: DUNN, Tracy
CC REGISTRATION NUMBER: 34,587
CC REFERENCE/DOCKET NUMBER: AVIR-5032
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 650.919.6637
CC TELEFAX: 650.919.6610
CC TELEX:
CC INFORMATION FOR SEQ ID NO: 21:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 642 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: Protein
SQ SEQUENCE 642 AA; 70848 MW; 2000477 CN;

Query Match 58.7%; Score 61; DB 13; Length 642;
Best Local Similarity 63.6%; Pred. No. 9.32e+01;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 404 LAIWQRTYNDT 414
||| |||
QY 3 LAVWKITYKDT 13

Search completed: Tue Aug 17 16:05:59 1999
Job time : 60 secs.

This Page Blank (uspto)

[WQSRHA]

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Aug 17 16:00:52 1999; MasPar time 6.53 Seconds
92.683 Million cell updates/sec

Tabular output not generated.

Title: >US-09-049-847-3
Description: (1-12) from US09049847.pep
Perfect Score: 82
Sequence: 1 GQIGNDPNRDIL 12

Scoring table: PAM 150
Gap 15

Searched: 411786 segs, 50406085 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-pending

1:P9 2:U60 3:U7 4:U80 5:U81 6:U82 7:U83 8:U84A 9:U84B
10:U85 11:U86 12:U87 13:U88 14:U89 15:U90 16:U91 17:U92
18:U93 19:NEWP 20:NEWU6 21:NEWU8 22:NEWU9

Statistics: Mean 17.684; Variance 44.259; scale 0.400

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	82	100.0	12	1	PCT-US98-0 Sequence 46, Applicatio	5.59e-03
2	82	100.0	12	15	US-09-049- Sequence 3, Applicatio	5.59e-03
3	82	100.0	12	18	US-09-321- Sequence 46, Applicatio	5.59e-03
4	82	100.0	12	15	US-09-003- Sequence 7, Applicatio	5.59e-03
5	82	100.0	13	12	US-08-788- Sequence 6, Applicatio	5.59e-03
6	82	100.0	13	9	US-08-485- Sequence 6, Applicatio	5.59e-03
7	82	100.0	13	9	US-08-485- Sequence 6, Applicatio	5.59e-03
8	82	100.0	13	7	US-08-305- Sequence 22, Applicatio	5.59e-03
9	82	100.0	14	5	US-08-121- Sequence 511, Applicatio	5.59e-03
10	82	100.0	14	15	US-09-082- Sequence 1, Applicatio	5.59e-03
11	82	100.0	1315	14	US-08-913- Sequence 19382, Applic	4.80e+01
12	53	64.6	373	2	US-09-096- Sequence 20, Applicatio	1.14e+02
13	53	64.6	373	17	US-09-248- Sequence 11, Applicatio	1.14e+02
14	50	61.0	20	7	US-08-374- Sequence 18, Applicatio	1.14e+02
15	50	61.0	432	7	US-08-374- Sequence 32, Applicatio	1.14e+02
16	50	61.0	504	4	US-08-089- Sequence 1, Applicatio	1.14e+02
17	50	61.0	521	10	US-08-569- Sequence 17, Applicatio	1.14e+02
18	50	61.0	521	17	US-09-262- Sequence 1, Applicatio	1.14e+02
19	50	61.0	522	7	US-08-310- Sequence 17, Applicatio	1.14e+02
20	50	61.0	530	10	US-08-569- Sequence 35, Applicatio	1.14e+02
21	50	61.0	530	17	US-09-262- Sequence 35, Applicatio	1.14e+02

22	50	61.0	635	17	US-09-216- Sequence 2, Applicatio	1.14e+02
23	50	61.0	3052	10	US-08-569- Sequence 8, Applicatio	1.14e+02
24	50	61.0	3052	17	US-09-282- Sequence 26, Applicati	1.14e+02
25	49	59.8	164	2	US-60-096- Sequence 17431, Applic	1.52e+02
26	49	59.8	164	17	US-09-248- Sequence 17431, Applic	1.52e+02
27	49	59.8	175	2	US-60-142- Sequence 424, Applicat	1.52e+02
28	49	59.8	175	2	US-60-141- Sequence 438, Applicat	1.52e+02
29	49	59.8	190	2	US-60-140- Sequence 722, Applicat	1.52e+02
30	49	59.8	220	14	US-08-983- Sequence 5, Applicatio	1.52e+02
31	49	59.8	356	20	US-60-143- Sequence 675, Applicat	1.52e+02
32	49	59.8	356	20	US-60-143- Sequence 551, Applicat	1.52e+02
33	49	59.8	584	14	US-08-983- Sequence 2, Applicatio	1.52e+02
34	49	59.8	841	15	US-09-092- Sequence 10, Applicatio	1.52e+02
35	49	59.8	1208	16	US-09-199- Sequence 1, Applicatio	1.52e+02
36	49	59.8	1218	11	US-08-611- Sequence 5, Applicatio	1.52e+02
37	49	59.8	1218	16	US-08-195- Sequence 5, Applicatio	1.52e+02
38	49	59.8	1346	11	US-08-635- Sequence 2, Applicatio	1.52e+02
39	49	59.8	1346	11	US-08-978- Sequence 2, Applicatio	1.52e+02
40	49	59.8	1346	1	PCT-US97-0 Sequence 2, Applicatio	1.52e+02
41	49	59.8	1346	11	US-08-665- Sequence 2, Applicatio	1.52e+02
42	49	59.8	1596	14	US-08-978- Sequence 4, Applicatio	1.52e+02
43	49	59.8	1596	1	PCT-US97-0 Sequence 4, Applicatio	1.52e+02
44	49	59.8	1596	11	US-08-665- Sequence 4, Applicatio	1.52e+02
45	48	58.5	57	1	PCT-US98-1 Sequence 467, Applicat	2.01e-02

ALIGNMENTS

RESULT 1

ID PCT-US98-01499-46 STANDARD; PRT; 12 AA.

XX AC xxxxxx

DE Sequence 46, Application PC/TUS9801499

XX Sequence 46, Application PC/TUS9801499

CC GENERAL INFORMATION:

CC APPLICANT: Pangaea, Inc.

CC TITLE OF INVENTION: MICROPARTICLES FOR DELIVERY

CC OF INVENTION: OF NUCLEIC ACID

CC NUMBER OF SEQUENCES: 107

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: Fish & Richardson, P.C.

CC STREET: 225 Franklin Street

CC CITY: Boston

CC STATE: MA

CC COUNTRY: US

CC ZIP: 02110-2804

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Diskette

CC COMPUTER: IBM Compatible

CC OPERATING SYSTEM: Windows95

CC SOFTWARE: FastSeq for Windows Version 2.0

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: PCT/US98/01499

CC FILING DATE: 22-JAN-1998

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: 08/787,547

CC FILING DATE: 22-JAN-1997

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Fraser, Janis K.

CC REGISTRATION NUMBER: 34,819

CC REFERENCE/DOCKET NUMBER: 08191/003WO1

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: 617-542-5070

CC TELEFAX: 617-542-8906

CC TELEX: 200154

CC INFORMATION FOR SEQ ID NO: 46:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 12 amino acids

CC TYPE: amino acid

```
CC      TOPOLOGY: linear
CC      MOLECULE TYPE: peptide
SQ      SEQUENCE 12 AA; 1311 MW; 550 CN;

Query Match      100.0%; Score 82; DB 1; Length 12;
Best Local Similarity 100.0%; Pred. No. 5.59e-03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      1 GQIGNDPNRDIL 12
QY      1 GQIGNDPNRDIL 12

RESULT      2
ID      US-09-049-847-3      STANDARD;      PRT;      12 AA.
XX
XX      xxxxxx
XX
DT
DE
DE      Sequence 3, Application US/09049847
CC
CC      Sequence 3, Application US/09049847
CC      GENERAL INFORMATION:
CC      APPLICANT: Bay, Sylvie
CC      APPLICANT: Cantacuzene, Daniele
CC      APPLICANT: Leclerc, Claude
CC      APPLICANT: Lo-Man, Richard
CC      TITLE OF INVENTION: Multiple antigen alycopeptide carbohydrate, vaccine
CC      FILE REFERENCE: 102.166A
CC      CURRENT APPLICATION NUMBER: US/09/049,847
CC      CURRENT FILING DATE: 1998-03-27
CC      EARLIER APPLICATION NUMBER: 60/041,726
CC      EARLIER FILING DATE: 1997-03-27
CC      NUMBER OF SEQ ID NOS: 4
CC      SOFTWARE: PatentIn Ver. 2.0
CC      SEQ ID NO 3
CC      LENGTH: 12
CC      TYPE: PRT
CC      ORGANISM: Clostridium tetani
SQ      SEQUENCE 12 AA; 1311 MW; 550 CN;

Query Match      100.0%; Score 82; DB 15; Length 12;
Best Local Similarity 100.0%; Pred. No. 5.59e-03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      1 GQIGNDPNRDIL 12
QY      1 GQIGNDPNRDIL 12

RESULT      3
ID      US-09-321-346-46      STANDARD;      PRT;      12 AA.
XX
XX      xxxxxx
XX
DT
DE
DE      Sequence 46, Application US/09321346
CC
CC      Sequence 46, Application US/09321346
CC      GENERAL INFORMATION:
CC      APPLICANT: Lunsford, Lynn B.
CC      APPLICANT: Putnam, David
CC      APPLICANT: Hedley, Mary Lynn
CC      TITLE OF INVENTION: MICROPARTICLES FOR DELIVERY OF NUCLEIC ACID
CC      FILE REFERENCE: 08191/014001
CC      CURRENT APPLICATION NUMBER: US/09/321,346
CC      CURRENT FILING DATE: 1999-05-27
CC      EARLIER APPLICATION NUMBER: US 09/266,463
CC      EARLIER FILING DATE: 1999-03-11
CC      NUMBER OF SEQ ID NOS: 114
```

```
CC      SOFTWARE: FastSEQ for Windows Version 3.0
CC      SEQ ID NO 46
CC      LENGTH: 12
CC      TYPE: PRT
CC      ORGANISM: Clostridium tetani
SQ      SEQUENCE 12 AA; 1311 MW; 550 CN;

Query Match      100.0%; Score 82; DB 18; Length 12;
Best Local Similarity 100.0%; Pred. No. 5.59e-03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      1 GQIGNDPNRDIL 12
QY      1 GQIGNDPNRDIL 12

RESULT      4
ID      US-09-003-253-46      STANDARD;      PRT;      12 AA.
XX
XX      xxxxxx
XX
DT
DE
DE      Sequence 46, Application US/09003253
CC
CC      Sequence 46, Application US/09003253
CC      GENERAL INFORMATION:
CC      APPLICANT: Hedley, Mary Lynne
CC      APPLICANT: Curley, Joanne M.
CC      APPLICANT: Langer, Robert S.
CC      TITLE OF INVENTION: MICROPARTICLES FOR DELIVERY OF NUCLEIC ACID
CC      NUMBER OF SEQUENCES: 108
CC      CORRESPONDENCE ADDRESS:
CC      ADDRESSEE: Fish & Richardson P.C.
CC      STREET: 225 Franklin Street
CC      CITY: Boston
CC      STATE: MA
CC      COUNTRY: USA
CC      ZIP: 02110-2804
CC      COMPUTER READABLE FORM:
CC      MEDIUM TYPE: Diskette
CC      COMPUTER: IBM Compatible
CC      OPERATING SYSTEM: Windows 95
CC      SOFTWARE: FastSEQ for Windows Version 2.0b
CC      CURRENT APPLICATION DATA:
CC      APPLICATION NUMBER: US/09/003,253
CC      FILING DATE: 06-JAN-1998
CC      CLASSIFICATION: 514
CC      PRIOR APPLICATION DATA:
CC      APPLICATION NUMBER: 50/035,983
CC      FILING DATE: 22-JAN-1997
CC      ATTORNEY/AGENT INFORMATION:
CC      NAME: Fraser, Ph.D., J.D., Janis K.
CC      REGISTRATION NUMBER: 34,819
CC      REFERENCE/DOCKET NUMBER: 08191/003002
CC      TELEPHONE: 617/542-5070
CC      TELEFAX: 617/542-8906
CC      TELEX: 200154
CC      INFORMATION FOR SEQ ID NO: 46:
CC      SEQUENCE CHARACTERISTICS:
CC      LENGTH: 12 amino acids
CC      TYPE: amino acid
CC      TOPOLOGY: linear
CC      MOLECULE TYPE: peptide
SQ      SEQUENCE 12 AA; 1311 MW; 550 CN;

Query Match      100.0%; Score 82; DB 15; Length 12;
Best Local Similarity 100.0%; Pred. No. 5.59e-03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      1 GQIGNDPNRDIL 12
QY      1 GQIGNDPNRDIL 12
```

```
0; Gaps 0;
```

CC APPLICANT: Grey, Howard M.
CC APPLICANT: Sidney, John
CC APPLICANT: Alexander, Jeffery L.
CC APPLICANT: Del Guercio, Marie-France
CC TITLE OF INVENTION: Alteration Of Immune Response Using Pan
CC TITLE OF INVENTION: DR-Binding Peptides
CC NUMBER OF SEQUENCES: 31
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Townsend and Townsend and Crew LLP
CC STREET: Two Embarcadero Center, Eighth Floor
CC CITY: San Francisco
CC STATE: California
CC COUNTRY: USA
CC ZIP: 94111-3834
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/485,218A
CC FILING DATE: 07-JUN-1995
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/305,871
CC FILING DATE: 14-SEP-1994
CC ATTORNEY/AGENT INFORMATION:
CC APPLICATION NUMBER: US 08/121,101
CC FILING DATE: 14-SEP-1993
CC NAME: Bastian, Kevin L.
CC REGISTRATION NUMBER: 34,774
CC REFERENCE/DOCKET NUMBER: 018623-006220US
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (415) 576-0200
CC TELEFAX: (415) 576-0300
CC INFORMATION FOR SEQ ID NO: 6:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 13 amino acids
CC TYPE: amino acid
CC STRANDEDNESS:
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 13 AA; 1426 MW; 637 CN;
SQ
Query Match 100.0%; Score 82; DB 9; Length 13;
Best Local Similarity 100.0%; Pred. No. 5.59e-03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 2 GQIGNDPNRDIL 13
QY 1 GQIGNDPNRDIL 12
RESULT 8
ID US-08-305-871-6 STANDARD; PRT; 13 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 6, Application US/08305871
XX
CC Sequence 6, Application US/08305871
CC GENERAL INFORMATION:
CC APPLICANT: Sette, Alessandro
CC APPLICANT: Gaeta, Federico
CC APPLICANT: Grey, Howard M.
CC APPLICANT: Sidney, John
CC APPLICANT: Alexander, Jeffery L.
CC TITLE OF INVENTION: Alteration Of Immune Response Using Pan
CC TITLE OF INVENTION: DR-Binding Peptides
CC NUMBER OF SEQUENCES: 17

CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Townsend and Townsend and Crew LLP
CC STREET: Two Embarcadero Center, Eighth Floor
CC CITY: San Francisco
CC STATE: California
CC COUNTRY: USA
CC ZIP: 94111-3834
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/305,871
CC FILING DATE: 14-SEP-1994
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/121,101
CC FILING DATE: 14-SEP-1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Bastian, Kevin L.
CC REGISTRATION NUMBER: 34,774
CC REFERENCE/DOCKET NUMBER: 14137-62-1
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (415) 576-0200
CC TELEFAX: (415) 576-0300
CC INFORMATION FOR SEQ ID NO: 6:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 13 amino acids
CC TYPE: amino acid
CC STRANDEDNESS:
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 13 AA; 1426 MW; 637 CN;
SQ
Query Match 100.0%; Score 82; DB 7; Length 13;
Best Local Similarity 100.0%; Pred. No. 5.59e-03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 2 GQIGNDPNRDIL 13
QY 1 GQIGNDPNRDIL 12
RESULT 9
ID US-08-121-101-22 STANDARD; PRT; 14 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 22, Application US/08121101
XX
CC Sequence 22, Application US/08121101
CC GENERAL INFORMATION:
CC APPLICANT: Sette, Alessandro
CC APPLICANT: Gaeta, Federico
CC APPLICANT: Grey, Howard M.
CC APPLICANT: Sidney, John
CC TITLE OF INVENTION: ALTERATION OF IMMUNE RESPONSE USING PAN
CC TITLE OF INVENTION: DR-BINDING PEPTIDES
CC NUMBER OF SEQUENCES: 25
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Townsend and Townsend Khourie and Crew
CC STREET: Steuart Street Tower, One Market Plaza
CC CITY: San Francisco
CC STATE: California
CC COUNTRY: US
CC ZIP: 94105-1493
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/121.101
FILING DATE: 14-SEP-1993
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Bastian, Kevin L.
REGISTRATION NUMBER: 34,774
REFERENCE/DOCKET NUMBER: 14137-62
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
SQ SEQUENCE 14AA: 1589 MW: 743 CN:

```

XX
AC      xxxxxxx
XX
XX
DT
XX
DE      Sequence 1, Application US/08913880
XX
CC      Sequence 1, Application US/08913880
CC      GENERAL INFORMATION:
CC      APPLICANT: MATSUDA, Morihiro
CC      TITLE OF INVENTION: TETANUS TOXIN FUNCTIONAL FRAGMENT ANTI GEN AND TETANUS
CC      FILE OF INVENTION: VACCINE
CC      FILE REFERENCE: 216-380P
CC      CURRENT APPLICATION NUMBER: US/08/913,880
CC      CURRENT FILING DATE: 1997-09-24
CC      NUMBER OF SEQ ID NOS: 9
CC      SOFTWARE: PatentIn Ver. 2.0
CC      SEQ ID NO 1
CC      LENGTH: 1315
CC      TYPE: PRT
CC      ORGANISM: Clostridium tetani
CC      SEQUENCE 1315 AA; 150681 MW; 9110516 CN;

Query Match      100.0%; Score 82; DB 14; Length 1315;
Best Local Similarity 100.0%; Pred. No. 5,59e-03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1273 GQIGNDFNRDIL 1284
      |||||
QY 1 GQIGNDFNRDIL 12

```

[illegible]

XX Sequence 19382, Application US/09248796
CC GENERAL INFORMATION:
CC APPLICANT: Keith Weinstock et al
CC TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALB
CC FILE REFERENCE: 107196.132
CC CURRENT APPLICATION NUMBER: US/09/248,796
CC CURRENT FILING DATE: 1999-02-12
CC NUMBER OF SEQ ID NOS: 28206
CC SEQ ID NO 19382
CC LENGTH: 373
CC TYPE: PRT
CC ORGANISM: Candida albicans
SQ SEQUENCE 373 AA; 41473 MW; 672994 CN;

Query Match 64.6%; Score 53; DB 17; Length 373;
Best Local Similarity 58.3%; Pred. No. 4.80e+01;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 183 SOIGNDPGRVL 194
QY 1 GOIGNDPNRDIL 12

RESULT 14
ID US-08-374-888A-20 STANDARD; PRT; 20 AA.
XX xxxxxx
AC
XX
DT
XX
DE Sequence 20, Application US/08374888A
XX
CC Sequence 20, Application US/08374888A
CC GENERAL INFORMATION:
CC APPLICANT: Hall, Linda M.
CC APPLICANT: Ren, Dejian
CC APPLICANT: Zheng, Wei
CC APPLICANT: Dubald, Manuel Marcel Paul
CC TITLE OF INVENTION: Genes Encoding an Insect Calcium Channel
CC NUMBER OF SEQUENCES: 101
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, LLP
CC STREET: 699 Prince Street
CC CITY: Alexandria
CC STATE: VA
CC COUNTRY: USA
CC ZIP: 22314-3187
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/374,888A
CC FILING DATE: 19-JAN-1995
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: McGowan, Malcolm M.
CC REGISTRATION NUMBER: 39,300
CC REFERENCE/DOCKET NUMBER: 022650-263
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 703-836-2021
CC TELEFAX: 703-836-2021
CC INFORMATION FOR SEQ ID NO: 20:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 20 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
SQ SEQUENCE 20 AA; 1952 MW; 1883 CN;

Query Match 61.0%; Score 50; DB 7; Length 432;
Best Local Similarity 50.0%; Pred. No. 1.14e+02;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 2 GOAGAEPSRDSM 13
QY 1 GOIGNDPNRDIL 12

RESULT 15
ID US-08-374-888-11 STANDARD; PRT; 432 AA.
XX xxxxxx
AC
XX
DT
XX
DE Sequence 11, Application US/08374888
XX
CC Sequence 11, Application US/08374888
CC GENERAL INFORMATION:
CC APPLICANT: HALL, Linda M
CC APPLICANT: REN, Dejian
CC APPLICANT: ZHENG, Wei
CC APPLICANT: DUBALD, Manuel Marcel Paul
CC TITLE OF INVENTION: GENES ENCODING AN INSECT CALCIUM CHANNEL
CC NUMBER OF SEQUENCES: 55
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Burns, Doane, Swecker & Mathis
CC STREET: P.O. Box 1404
CC CITY: Alexandria
CC STATE: Virginia
CC COUNTRY: United States
CC ZIP: 22313-1404
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/374,888
CC FILING DATE: 19-JAN-1995
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: McGowan, Malcolm K
CC REGISTRATION NUMBER: P39,300
CC REFERENCE/DOCKET NUMBER: 022650-263
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (703) 836-6620
CC TELEFAX: (703) 836-2021
CC INFORMATION FOR SEQ ID NO: 11:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 432 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: DNA (genomic)
SQ SEQUENCE 432 AA; 47886 MW; 945243 CN;

Query Match 61.0%; Score 50; DB 7; Length 432;
Best Local Similarity 50.0%; Pred. No. 1.14e+02;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 160 GOAGAEPSRDSM 171
QY 1 GOIGNDPNRDIL 12

Search completed: Tue Aug 17 16:01:46 1999
Job time : 54 secs.

Query Match 61.0%; Score 50; DB 7; Length 20;
Best Local Similarity 50.0%; Pred. No. 1.14e+02;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 2 GOAGAEPSRDSM 13
QY 1 GOIGNDPNRDIL 12

RESULT 15
ID US-08-374-888-11 STANDARD; PRT; 432 AA.
XX xxxxxx
AC
XX
DT
XX
DE Sequence 11, Application US/08374888
XX
CC Sequence 11, Application US/08374888
CC GENERAL INFORMATION:
CC APPLICANT: HALL, Linda M
CC APPLICANT: REN, Dejian
CC APPLICANT: ZHENG, Wei
CC APPLICANT: DUBALD, Manuel Marcel Paul
CC TITLE OF INVENTION: GENES ENCODING AN INSECT CALCIUM CHANNEL
CC NUMBER OF SEQUENCES: 55
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Burns, Doane, Swecker & Mathis
CC STREET: P.O. Box 1404
CC CITY: Alexandria
CC STATE: Virginia
CC COUNTRY: United States
CC ZIP: 22313-1404
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/374,888
CC FILING DATE: 19-JAN-1995
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: McGowan, Malcolm K
CC REGISTRATION NUMBER: P39,300
CC REFERENCE/DOCKET NUMBER: 022650-263
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (703) 836-6620
CC TELEFAX: (703) 836-2021
CC INFORMATION FOR SEQ ID NO: 11:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 432 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: DNA (genomic)
SQ SEQUENCE 432 AA; 47886 MW; 945243 CN;

Query Match 61.0%; Score 50; DB 7; Length 432;
Best Local Similarity 50.0%; Pred. No. 1.14e+02;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 160 GOAGAEPSRDSM 171
QY 1 GOIGNDPNRDIL 12

Search completed: Tue Aug 17 16:01:46 1999
Job time : 54 secs.

W P S R L H

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein ~ protein database search, using Smith-Waterman algorithm

Run on: Tue Aug 17 16:10:27 1999; MasPar time 4.20 Seconds
Tabular output not generated. 106.318 Million cell updates/sec

Title: >US-09-049-847-2
Description: (1-21) from US09049847.pep
Perfect Score: 158
Sequence: 1 FNNFTVSEWLRVPKVSASHLE 21

Scoring table: PAM 150
Gap 15

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-genesec35

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 21.709; Variance 73.372; scale 0.296

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	158	100.0	21	38	T-cell epitope peptid	4.34e-09
2	158	100.0	21	2	Immunogenic conjugate	4.34e-09
3	158	100.0	21	36	Tetanus toxin fragmen	4.34e-09
4	158	100.0	21	20	Tetanus toxoid protei	4.34e-09
5	158	100.0	21	28	Broad range helper T	4.34e-09
6	158	100.0	21	17	T-cell antigen Tr3 pe	4.34e-09
7	158	100.0	21	37	Tetanus toxoid epitop	4.34e-09
8	158	100.0	32	13	LHRH-containing immu	4.34e-09
9	158	100.0	63	3	Immunogenic branched	4.34e-09
10	158	100.0	64	3	Immunogenic branched	4.34e-09
11	158	100.0	65	3	Immunogenic branched	4.34e-09
12	158	100.0	65	3	Immunogenic branched	4.34e-09
13	158	100.0	77	3	Immunogenic branched	4.34e-09
14	158	100.0	452	2	Tetanus toxin fragmen	4.34e-09
15	158	100.0	573	2	Portion of B fragment	4.34e-09
16	158	100.0	518	33	SOD-1/TTC hybrid prot	4.34e-09

17	152	96.2	22	16	R82574	Tetanus toxin helper	2.09e-08
18	152	96.2	22	13	R62593	Helper T cell epitope	2.09e-08
19	152	96.2	22	19	W05600	Tetanus toxin helper	2.09e-08
20	152	96.2	32	16	R82597	IGF CH4 region contg.	2.09e-08
21	152	96.2	34	16	R83562	IGF CH4 region contg.	2.09e-08
22	135	85.4	20	35	W71322	Universal helper T-ce	1.71e-06
23	132	83.5	20	34	W69279	Tetanus toxin fragmen	3.88e-06
24	129	81.6	21	27	W35438	T-cell stimulatory pe	7.92e-06
25	116	73.4	14	35	W78832	Tetanus toxoid protei	2.12e-04
26	114	72.2	15	35	W78833	Tetanus toxoid protei	3.50e-04
27	114	72.2	22	20	W03002	Carrier peptide for a	3.73e-02
28	95	60.1	473	35	W68400	Clostridium botulinum	6.03e-02
29	93	58.9	144	20	W09015	Immunogenic type F bo	6.03e-02
30	93	58.9	431	20	W09014	Immunogenic type F bo	6.03e-02
31	93	58.9	448	55	W68399	Clostridium botulinum	6.03e-02
32	93	58.9	472	35	W68393	Clostridium botulinum	6.03e-02
33	93	58.9	472	35	W68394	Clostridium botulinum	6.03e-02
34	93	58.9	1159	30	W56017	Recombinant botulinum	6.03e-02
35	93	58.9	1291	35	W68392	Clostridium botulinum	6.03e-02
36	90	57.0	438	17	R95008	Type A neurotoxin C f	1.24e-01
37	90	57.0	438	55	W68389	Clostridium botulinum	1.24e-01
38	90	57.0	445	35	W68391	Clostridium botulinum	1.24e-01
39	90	57.0	462	17	R95009	Type A neurotoxin C f	1.24e-01
40	90	57.0	462	35	W68390	Clostridium botulinum	1.24e-01
41	90	57.0	1296	17	R95010	Clostridium botulinum	1.24e-01
42	88	55.7	451	35	W68395	Clostridium botulinum	1.99e-01
43	88	55.7	452	35	W68396	Clostridium botulinum	1.99e-01
44	84	53.2	260	12	R56546	Partial sequence of n	5.10e-01
45	84	53.2	464	2	R11834	Non-structural protei	5.10e-01

ALIGNMENTS

RESULT 1
ID W67579 standard; peptide: 21 AA.
AC W67579;
DT 02-MAR-1999 (first entry)
DE T-cell epitope peptide #5 for chimeric fimbria/T-cell epitope peptide.
KW Chimeric; non-typable Haemophilus influenzae; fimbria; T-cell epitope;
KW immunogenic composition; immune response.
OS Synthetic.
PN US5843464-A.
PD 01-DEC-1998.
PF 02-JUN-1995; 460502.
PR 02-JUN-1995; US-460502.
PI (OHIS) UNIV OHIO STATE.
PI Bakaletz LO, Kaumaya PTP.
DR WPI; 99-044514/04.
PT Synthetic chimeric fimbria peptide - useful for vaccination against
PT non-typable Haemophilus influenzae
PS Disclosure; Column 4; 16pp; English.
CC The invention relates to the manufacture of a synthetic chimeric peptide
CC comprising a non-typable Haemophilus influenzae fimbria peptide fused via
CC a linker peptide to a T-cell epitope peptide. The chimeric peptide is
CC used in immunogenic compositions which induce an immune response against
CC non-typable Haemophilus influenzae. This sequence represents an example
CC of a T-cell epitope peptide used to generate the chimeric peptide.
SQ Sequence 21 AA;

Query Match 100.0%; Score 158; DB 38; Length 21;
Best Local Similarity 100.0%; Pred No. 4.34e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 fnnftvswlrvpkvsashle 21
Qy 1 FNNFTVSEWLRVPKVSASHLE 21
|||||

RESULT 2
ID R11896 standard; peptide: 21 AA.
AC R11896;
DT 19-JUL-1991 (first entry)
DE Immunogenic conjugate constituent peptide, TT3.

KW Malaria vaccine; major histocompatibility complex.
 OS Tetanus toxin.
 PH Key Location/Qualifiers
 FT peptide 1..14
 FT /label= active fragment (claimed)
 PN EP-427347-A.
 PD 15-MAY-1991.
 PE 07-NOV-1990; 202948.
 PR 10-NOV-1989; IT-022355.
 PA (ENIE) ENRICERCHÉ SPA.
 PI Bianchi E, Pessi A, Corradin G;
 DR WPI; 91-141874/20.
 PT Synthetic peptide(s) used as universal carriers - for preparing
 PT immunogenic conjugates used as vaccines against Plasmodium
 PT falciparum
 PS Claim 1; page 13; 16pp; English.
 CC This peptide corresponds to residues 947-967 of Tetanus toxin. It can
 CC be used as a universal carrier for the prep. of an immunogenic
 CC conjugate. It is covalently bound to a peptide or polysaccharide
 CC haptén derived from a pathogen. This conjugate can be used as
 CC a vaccine for malaria. This peptide is recognised by different T-
 CC helper cell clones in association with alleles of the human MHC.
 CC It contains 2 epitopes: (a) 933-967, recognised by DR5-restricted
 CC clones; and (b) 947-960, recognised by all other DR and DP-
 CC restricted clones.
 CC Sequence 21 AA;
 SQ

Query Match 100.0%; Score 158; DB 2; Length 21;
 Best Local Similarity 100.0%; Pred. No. 4.34e-09;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 fnftvfwlrvpkvsashle 21
 QY 1 FNNFTVFWLVRPKVSASHLE 21

RESULT 3
 ID W67034 standard; peptide: 21 AA.
 AC W67034;
 DT 15-DEC-1998 (first entry)
 DE Tetanus toxin fragment (residues 947-967).
 KW Tetanus toxin; vaccine; antibody; carbohydrate peptide conjugate;
 KW dendrimeric poly-L-lysine; epitope; tumour.
 OS Clostridium tetani.
 PN W09843677-A1.
 PD 08-OCT-1998.
 PF 27-MAR-1998; E01922.
 PR (INSP) INST PASTEUR.
 PA Bay S, Cantacuzene D, Leclerc C, Lo-man R;
 PI WPI; 98-557071/47.
 DR Carbohydrate peptide conjugate used as vaccine - comprises carrier
 PT with dendrimeric poly-L-lysine enabling multiple epitopes to be
 PT covalently attached
 PS Disclosure; Page 13; 55pp; English.
 CC The invention relates to a new carbohydrate peptide conjugate, which
 CC comprises a carrier with a dendrimeric poly-L-lysine enabling multiple
 CC epitopes to be covalently attached to it. Also claimed are: (1) an
 CC antibody purified from biological fluid or cells of organisms
 CC administered with the carbohydrate peptide conjugate, and (2) a diagnosis
 CC kit comprising antigen-specific antibodies elicited by immunisation with
 CC the carbohydrate peptide conjugate. The peptide conjugate, antibody and
 CC diagnosis kit are used to provide pharmaceutical compositions and
 CC vaccines against tumours. These can be used to support an immune response
 CC against viral infections caused by hepatitis virus, HIV or cytomegalovirus.
 CC They can be used to enhance immune responses, especially B- and T-
 CC cell responses, of humans and animals against bacterial infections. The
 CC carbohydrate peptide conjugate stimulates the antibody and T-cell
 CC response without stimulating undesired immune responses. The composition
 CC is capable of increasing the survival of tumour bearing humans and
 CC animals. The present sequence corresponds to residues 947-967 of tetanus
 CC toxin. The synthetic peptide corresponding to this sequence may be used
 CC as an epitope in a carbohydrate peptide conjugate.

SQ Sequence 21 AA;
 Query Match 100.0%; Score 158; DB 36; Length 21;
 Best Local Similarity 100.0%; Pred. No. 4.34e-09;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 fnftvfwlrvpkvsashle 21
 QY 1 FNNFTVFWLVRPKVSASHLE 21

RESULT 4
 ID W06130 standard; Peptide: 21 AA.
 AC W06130;
 DT 07-FEB-1997 (first entry)
 DE Tetanus toxoid protein T-cell epitope.
 KW Cholesteryl ester transfer protein; CERP; antigen; vaccine;
 KW cardiovascular disease; atherosclerosis; tetanus toxoid;
 KW T-cell epitope.
 OS Clostridium tetani.
 PN W09634888-A1.
 PD 07-NOV-1996.
 PF 01-MAY-1996; U06147.
 PR 01-MAY-1995; US-432483.
 PA (TCEL-) T CELL SCI INC.
 PI Rittershaus CW, Thomas LJ;
 DR WPI; 96-506103/50.
 PT Cholesteryl ester transfer protein B cell epitope linked to T cell
 PT epitope - used to generate vaccine to regulate CERP activity for
 PT decreasing the risk of developing a cardiovascular disease e.g.
 PT atherosclerosis
 PS Claim 11; Page 43; 72pp; English.
 CC A helper T-cell epitope (W06130) comprises amino acids 947-967
 CC of tetanus toxoid protein. It can be utilised in novel peptide
 CC vaccines (see also W06129, W06132) also including B-cell
 CC epitope(s) from human or rabbit cholesteryl ester transfer
 CC protein (CERP) to elicit an immune response against endogenous
 CC CERP activity, thereby treating or preventing a cardiovascular
 CC disease, such as atherosclerosis.
 CC Sequence 21 AA;
 SQ

Query Match 100.0%; Score 158; DB 20; Length 21;
 Best Local Similarity 100.0%; Pred. No. 4.34e-09;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 fnftvfwlrvpkvsashle 21
 QY 1 FNNFTVFWLVRPKVSASHLE 21

RESULT 5
 ID W46449 standard; Peptide: 21 AA.
 AC W46449;
 DT 18-MAY-1998 (first entry)
 DE Broad range helper T cell epitope from the tetanus toxoid protein.
 KW Cholesteryl ester transfer protein; CERP; cholesteryl ester;
 KW high density lipoprotein; HDL; very low density lipoprotein; VLDL;
 KW low density lipoprotein; LDL; T cell epitope; antibody;
 KW DNA plasmid-based vaccine; broad range helper T cell epitope;
 KW treatment; cardiovascular disease.
 OS Clostridium tetani.
 PN W09741227-A1.
 PD 06-NOV-1997.
 PF 01-MAY-1997; U07294.
 PR 21-FEB-1997; US-802967.
 PR 01-MAY-1996; US-640713.
 PA (TCEL-) T CELL SCI INC.
 PI Thomas LJ;
 DR WPI; 97-549731/50.
 PT DNA plasmid-based vaccine encodes CERP B cell and helper T cell
 PT epitope(s) - used for elevating high density lipoprotein levels, and
 PT for treating cardiovascular disease
 PS Disclosure; Page 44; 67pp; English.

CC The present sequence represents a broad range helper T cell epitope
 CC of the tetanus protein. It can be used in DNA plasmid-based vaccines
 CC against cholesteryl ester transfer proteins (CETPs). CETPs mediate the
 CC transfer of cholesteryl esters from high density lipoprotein (HDL)
 CC to very low density lipoprotein (VLDL) and low density lipoprotein (LDL),
 CC and vice versa. An increased CETP activity produces an atherogenic
 CC lipoprotein profile and induces atherosclerosis. A DNA plasmid-based
 CC vaccine comprises sequences encoding at least one B cell epitope of CETP
 CC linked in frame with at least one segment encoding a broad range helper
 CC T cell epitope. The vaccines can be used to elevate the ratio of
 CC circulating HDL to circulating LDL, VLDL or total cholesterol in a human.
 CC It can also be used for decreasing the level of endogenous CETP activity
 CC in a human. The vaccine can be used to produce anti-CETP antibodies in
 CC vivo and for treating cardiovascular disease.
 SQ Sequence 21 AA;

Query Match 100.0%; Score 158; DB 28; Length 21;
 Best Local Similarity 100.0%; Pred. No. 4.34e-09;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 fnnftsvfwrpkvsashle 21
 QY 1 FNNFTSVFWRPKVSASHLE 21
 |||||

RESULT 6

ID R88397 standard; Peptide; 21 AA.

AC R88397; AC 12-JUN-1996 (first entry)
 DE T-cell antigen IT3 peptide.
 KW T-antigen; vaccine; antibody; T-cell; T-lymphocyte;
 KW alpha-helix; coiled-coil heterodimer; core peptide; subunit.
 OS Synthetic.

PN W09531480-A1.

PD 23-NOV-1995.

PF 18-MAY-1995; CA0293.

PR 18-MAY-1994; US-245507.

PA (SPIS-) SPI SYNTHETIC PEPTIDES INC.

PI Caccia PJ, Hodges RS, Houston ME, Irvin RT, Kay CM;
 PI Zhou NE;

DR WPI; 96-010880/01.

PT Heterodimeric polypeptide immunogen in coiled-coil configuration
 PT with different antigens on each sub:unit - useful in vaccines and
 PT for antibody prodn.

PS Claim 7; Page 62; 95pp; English.

CC This T-cell antigen IT3 peptide may be attached to a core peptide
 CC contained in one of the 2 subunits of an alpha-helical coiled-coil
 CC heterodimer. Each core peptide is comprised of terminal and
 CC internal AA repeat sequences. This peptide antigen is attached

CC to the core peptide through covalent linkages to certain AA of the
 CC internal repeats. The 2 subunits of the heterodimer are arranged
 CC in a stable alpha-helical coiled-coil configuration having a 1:1
 CC stoichiometry, and the peptide antigen is disposed toward the outer
 CC surfaces of the configuration. The heterodimer may be used as a
 CC synthetic vaccine (optionally multivalent) or to generate
 CC antibodies.

SQ Sequence 21 AA;

Query Match 100.0%; Score 158; DB 17; Length 21;
 Best Local Similarity 100.0%; Pred. No. 4.34e-09;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 fnnftsvfwrpkvsashle 21
 QY 1 FNNFTSVFWRPKVSASHLE 21
 |||||

RESULT 7

ID W73222 standard; Protein; 21 AA.

AC W73222;

DT 25-JAN-1999 (first entry)

DE Tetanus toxoid epitope.

KW Multispecific single chain antibody; antibody H22; tumour cell; therapy;

KW antibody-dependent cellular cytotoxicity; ADCC; HER 2/neu; infection;
 KW epidermal growth factor receptor; breast cancer; ovarian cancer.
 OS Synthetic.
 PN US5837243-A.
 PD 17-NOV-1998.
 PF 07-JUN-1996; 661052.
 PR 07-JUN-1995; US-661052.
 PR 07-JUN-1995; US-484172.
 PA (MEDA-) MEDAREX INC.
 PI Deo YM, Goldstein J, Graziano R, Somasundaram C;
 DR WPI; 99-023374/02.
 PT Specific killing of tumour cells - using a multi-specific molecule
 PT comprising an anti-Fc receptor antibody and a portion which binds to
 PT a target cell
 PS Example 7; Column 27; 57pp; English.
 CC This sequence represents a tetanus toxoid epitope and is recognised
 CC by the multispecific single chain antibody designated H22. The
 CC antibody can be used in the method of the invention for inducing
 CC antibody-dependent cellular cytotoxicity (ADCC) against a tumour cell
 CC which is characterised by overexpression of HER 2/neu or epidermal growth
 CC factor receptor (EGFR), comprises contacting the tumour cell with a
 CC multispecific protein molecule (preferably a single chain antibody)
 CC comprising: (a) an anti-Fc receptor antibody or an antigen binding
 CC fragment; (b) a portion which binds to HER 2/neu; and (c) a portion which
 CC binds to EGFR. The method can be used for treating cancers especially
 CC breast cancer or ovarian cancer. The multispecific antibody can also
 CC be administered prophylactically to vaccinate a subject against infection
 CC by a target cell.
 SQ Sequence 21 AA;

Query Match 100.0%; Score 158; DB 37; Length 21;
 Best Local Similarity 100.0%; Pred. No. 4.34e-09;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 fnnftsvfwrpkvsashle 21
 QY 1 FNNFTSVFWRPKVSASHLE 21
 |||||

RESULT 8

ID R62702 standard; peptide; 32 AA.

AC R62702;

DT 10-SEP-1995 (first entry)

DE LHRH-containing immunogenic peptide.

KW Helper T cell epitope; universal immune stimulator; invasin; hapten;

KW vaccine; LHRH; luteinising hormone releasing hormone; prostate;

KW androgen-dependent carcinoma; antitumour; infertility;

KW tetanus toxin.

OS Synthetic.

FH Key Location/Qualifiers

FT domain 1..22

FT domain /note= "tetanus toxin helper T cell epitope"

FT domain 23..32

FT domain /note= "LHRH hapten"

PN W09425060-A.

PD 10-NOV-1994.

PF 28-APR-1994; U04832.

PR 27-APR-1993; US-057166.

PR 14-APR-1994; US-229275.

PA (LADD/) LADD A E.

PA (WANG/) WANG C Y.

PA (ZAMB/) ZAMB T.

PI Ladd AE, Wang CY, Zamb T;

DR WPI; 94-357910/44.

PT Immunogenic luteinising hormone releasing hormone peptide(s) -

PT that suppress LHRH activity in males and females

PS Claim 8; Page 84; 213pp; English.

CC Synthetic immunogenic peptides are provided in which a universal immune

CC stimulator is linked to a peptide or protein hapten containing B cell

CC and/or cytotoxic T lymphocyte epitopes, giving a product which causes

CC potent immune responses to the coupled peptide or protein. The

CC stimulator consists of (A) a promiscuous helper T cell epitope (Th)

CC which elicits an immune response to the coupled peptide in members of

CC a heterogeneous population expressing diverse HLA phenotypes, and (B)
 CC an adjuvant peptide sequence from the invasin protein of *Yersinia*.
 CC Spacer amino acid sequences (e.g. Gly-Gly) can be provided between the
 CC invasin and Th domains and between the immune stimulator and hapten
 CC components. When the hapten is LHRH, then optionally the invasin domain
 CC can be omitted from the immune stimulator component.
 CC The present sequence represents an LHRH-containing, invasin-free
 CC immunogenic peptide as above which can be used as a potent vaccine for
 CC treating e.g. prostatic hyperplasia, androgen-dependent carcinoma,
 CC prostatic carcinoma, testicular carcinoma, endometriosis, benign
 CC uterine tumours, recurrent functional ovarian cysts, (severe) for
 CC premenstrual syndrome or oestrogen-dependent breast cancer, or for
 CC induction of infertility.
 SQ Sequence 32 AA;

Query Match 100.0%; Score 158; DB 13; Length 32;
 Best Local Similarity 100.0%; Pred. No. 4.34e-09;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 3 ffnftvswlrvpkvsashle 23
 |||||
 QY 1 ffnftvswlrvpkvsashle 21

RESULT 9
 ID R14263 standard; Peptide; 63 AA.
 AC R14263;
 DT 14-JAN-1992 (first entry)
 DE Immunogenic branched polypeptides for antimalarial vaccines.
 KW Immunogen; Plasmodium; malaria; lysine; immunosassay.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT region 25..38
 FT /label= I epitope
 FT region 39..59
 FT /label= T epitope
 FT modified_site 60
 FT /note= "epsilon amino substituted with the sequence
 (NANP)60YIKANSKFIGITENFTVSWLVRPKVSASHLE"
 FT modified_site 61
 FT /note= "epsilon amino substituted by Lys in which
 both alpha and epsilon amino groups are
 substituted with the sequence
 (NANP)60YIKANSKFIGITENFTVSWLVRPKVSASHLE"
 FT modified_site 62
 FT /note= "epsilon amino substituted by Lys in which
 both alpha and epsilon amino groups are
 substituted by Lys, both of the latter
 two Lys residues being substituted in each
 of their alpha and epsilon amino groups by
 (NANP)60YIKANSKFIGITENFTVSWLVRPKVSASHLE"
 CC The peptide is a specific example of highly generic
 CC immunogenic substituted lysines or polylysines having a number n
 CC (where n is 1-15) of L-lysine amino acid residues of alpha and
 CC epsilon amide linkage, where (n+1)/2 of the alpha amino groups
 CC and/or (n+1)/2 of the epsilon amino groups are substituted with
 CC polypeptides consisting of one or more plasmodial B epitopes
 CC covalently bound to one or more peptides with an amino acid sequence
 CC corresponding to that of a T epitope such as FNNFTVSWLVRPKVSASHLEA
 CC or OYIKANSKFIGITE
 CC The branched polypeptides can be used as immunogens for
 CC preparing genetically non-restricted antimalaria vaccines and for

CC determining anti-Plasmodium antibodies in blood, serum and blood-spot
 CC samples. Determination can be effected by ELISA.
 CC See also R14261-2, R14264-5 and R15436.
 SQ Sequence 63 AA;

Query Match 100.0%; Score 158; DB 3; Length 63;
 Best Local Similarity 100.0%; Pred. No. 4.34e-09;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 39 ffnftvswlrvpkvsashle 59
 |||||
 QY 1 ffnftvswlrvpkvsashle 21

RESULT 10
 ID R14261 standard; peptide; 64 AA.
 AC R14261;
 DT 14-JAN-1992 (first entry)
 DE Immunogenic branched polypeptides for antimalarial vaccines.
 KW Immunogen; Plasmodium; malaria; lysine; immunosassay.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT region 1..21
 FT /label= T epitope
 FT modified_site 62
 FT /note= "epsilon-amino substituted with the sequence
 FNNFTVSWLVRPKVSASHLE(NANP)10 "
 FT modified_site 63
 FT /note= "epsilon-amino group substituted with Lys in
 which both alpha and epsilon amino groups
 are substituted with the sequence
 FNNFTVSWLVRPKVSASHLE(NANP)10"
 PN EP-450715-A.
 PD 09-OCT-1991.
 PE 28-MAR-1991; 200727.
 PR 02-APR-1990; IT-019914.
 PA (ENIE) ENIRICERHE SPA.
 PI Pessi A, Bianchi E, Corradin G;
 DR WPI; 91-297504/41.
 PT New immunogenic branched polypeptide derivs. - used as antigens
 PT in enzyme immunoassays and as anti sporozoite vaccines against
 Plasmodium falciparum
 PS Claim 8; Page 15; 22pp; English.
 CC The peptide is a specific example of highly generic
 CC immunogenic substituted lysines or polylysines having a number n
 CC (where n is 1-15) of L-lysine amino acid residues of alpha and
 CC epsilon amide linkage, where (n+1)/2 of the alpha amino groups
 CC and/or (n+1)/2 of the epsilon amino groups are substituted with
 CC polypeptides consisting of one or more plasmodial B epitopes
 CC covalently bound to one or more peptides with an amino acid sequence
 CC corresponding to that of a T epitope such as FNNFTVSWLVRPKVSASHLEA
 CC or OYIKANSKFIGITE
 CC The branched polypeptides can be used as immunogens for
 CC preparing genetically non-restricted antimalaria vaccines and for
 CC determining anti-Plasmodium antibodies in blood, serum and blood-spot
 CC samples. Determination can be effected by ELISA.
 CC See also R14262 - R14265 and R15436.
 SQ Sequence 64 AA;

Query Match 100.0%; Score 158; DB 3; Length 64;
 Best Local Similarity 100.0%; Pred. No. 4.34e-09;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 1 ffnftvswlrvpkvsashle 21
 |||||
 QY 1 ffnftvswlrvpkvsashle 21
 RESULT 11
 ID R14265 standard; Peptide; 65 AA.
 AC R14265;
 DT 14-JAN-1992 (first entry)
 DE Immunogenic branched polypeptides for antimalarial vaccines.

KW Immunogen; Plasmodium; malaria; lysine; immunoassay.
 OS Synthetic.
 FH Key
 FT region Location/Qualifiers
 FT 1..21 /label= T epitope
 FT modified_site 63
 FT /note= "epsilon amino substituted with the sequence
 FT FNNFTVSFWLRVPKVSASHLE(NANP)10K"
 FT modified_site 64
 FT /note= "epsilon amino substituted with Lys in which
 FT alpha and epsilon amino groups are each
 FT substituted with the sequence
 FT FNNFTVSFWLRVPKVSASHLE(NANP)10K"
 FT
 PN EP-450715-A.
 PD 09-OCT-1991.
 PR 28-MAR-1991; 200727.
 PR 02-APR-1990; IT-019914.
 PA (ENTE) ENTRICERCH SPA.
 PI Pessi A, Bianchi E, Corradin G;
 DR WPI: 91-297504/41.
 FT New immunogenic branched polypeptide derivs. - used as antigens
 FT in enzyme immunoassays and as anti sporozoite vaccines against
 FT Plasmodium falciparum
 PS Claim 13; Page 16; 22pp; English.
 CC The peptide is a specific example of highly generic
 CC immunogenic substituted lysines or polylysines having a number n
 CC (where n is 1-15) of L-lysine amino acid residues of alpha and
 CC epsilon amide linkage, where (n+1)/2 of the alpha amino groups
 CC and/or (n+1)/2 of the epsilon amino groups are substituted with
 CC polypeptides consisting of one or more plasmodial B epitopes
 CC covalently bound to one or more peptides with an amino acid sequence
 CC corresponding to that of a T epitope such as FNNFTVSFWLRVPKVSASHLEA
 CC or QYKANSKFIGITE.
 CC The branched polypeptides can be used as immunogens for
 CC preparing genetically non-restricted antimalaria vaccines and for
 CC determining anti-Plasmodium antibodies in blood, serum and blood-spot
 CC samples. Determination can be effected by ELISA.
 CC See also R14261 - R14264, R14266 and R15436.
 CC Sequence 65 AA;
 SQ
 Query Match 100.0%; Score 158; DB 3; Length 65;
 Best Local Similarity 100.0%; Pred. No. 4.34e-09;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 1 fnnftvsfwlrvpkvsashle 21
 QY 1 FNNFTVSFWLRVPKVSASHLE 21
 RESULT 12
 ID R14262 standard; Peptide; 65 AA.
 AC R14262;
 DT 14-JAN-1992 (first entry)
 DE Immunogenic branched polypeptides for antimalarial vaccines.
 KW Immunogen; Plasmodium; malaria; lysine; immunoassay.
 OS Synthetic.
 FH Key
 FT region Location/Qualifiers
 FT 1..21 /label= T epitope
 FT modified_site 62
 FT /note= "epsilon amino substituted by the sequence
 FT VQGEESNDK"
 FT modified_site 63
 FT /note= "epsilon amino substituted by Lys in which the
 FT alpha amino is substituted with the sequence
 FT FNNFTVSFWLRVPKVSASHLE(NANP)10 and the epsilon
 FT amino is substituted with the sequence
 FT VQGEESNDK"
 FT modified_site 64
 FT /note= "epsilon amino substituted by Lys in which
 FT both the alpha and epsilon amino groups are
 FT substituted with further Lys residues, the
 FT latter two Lys residues each being substituted

FT On the alpha amino by FNNFTVSFWLRVPKVSASHL-
 FT E(NANP)10 and on the epsilon amino by the
 FT sequence VQGEESNDK"
 PN EP-450715-A.
 PD 09-OCT-1991.
 PR 28-MAR-1991; 200727.
 PR 02-APR-1990; IT-019914.
 PA (ENTE) ENTRICERCH SPA.
 PI Pessi A, Bianchi E, Corradin G;
 DR WPI: 91-297504/41.
 FT New immunogenic branched polypeptide derivs. - used as antigens
 FT in enzyme immunoassays and as anti sporozoite vaccines against
 FT Plasmodium falciparum
 PS Claim 9; Page 15; 22pp; English.
 CC The peptide is a specific example of highly generic
 CC immunogenic substituted lysines or polylysines having a number n
 CC (where n is 1-15) of L-lysine amino acid residues of alpha and
 CC epsilon amide linkage, where (n+1)/2 of the alpha amino groups
 CC and/or (n+1)/2 of the epsilon amino groups are substituted with
 CC polypeptides consisting of one or more plasmodial B epitopes
 CC covalently bound to one or more peptides with an amino acid sequence
 CC corresponding to that of a T epitope such as FNNFTVSFWLRVPKVSASHLEA
 CC or QYKANSKFIGITE.
 CC The branched polypeptides can be used as immunogens for
 CC preparing genetically non-restricted antimalaria vaccines and for
 CC determining anti-Plasmodium antibodies in blood, serum and blood-spot
 CC samples. Determination can be effected by ELISA.
 CC See also R14262 - R14265 and R15436.
 CC Sequence 65 AA;
 SQ
 Query Match 100.0%; Score 158; DB 3; Length 65;
 Best Local Similarity 100.0%; Pred. No. 4.34e-09;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 1 fnnftvsfwlrvpkvsashle 21
 QY 1 FNNFTVSFWLRVPKVSASHLE 21
 RESULT 13
 ID R14264 standard; Peptide; 77 AA.
 AC R14264;
 DT 14-JAN-1992 (first entry)
 DE Immunogenic branched polypeptides for antimalarial vaccines.
 KW Immunogen; Plasmodium; malaria; lysine; immunoassay.
 OS Synthetic.
 FH Key
 FT region Location/Qualifiers
 FT 13..33 /label= T epitope
 FT modified_site 75
 FT /note= "epsilon amino substituted with the sequence
 FT (NANP)3FNNFTVSFWLRVPKVSASHLE(NANP)10K"
 FT modified_site 76
 FT /note= "epsilon amino substituted with Lys in which
 FT alpha and epsilon amino groups are each
 FT substituted with the sequence
 FT (NANP)3FNNFTVSFWLRVPKVSASHLE(NANP)10K"
 FT
 PN EP-450715-A.
 PD 09-OCT-1991.
 PR 28-MAR-1991; 200727.
 PR 02-APR-1990; IT-019914.
 PA (ENTE) ENTRICERCH SPA.
 PI Pessi A, Bianchi E, Corradin G;
 DR WPI: 91-297504/41.
 FT New immunogenic branched polypeptide derivs. - used as antigens
 FT in enzyme immunoassays and as anti sporozoite vaccines against
 FT Plasmodium falciparum
 PS Claim 11; Page 16; 22pp; English.
 CC The peptide is a specific example of highly generic
 CC immunogenic substituted lysines or polylysines having a number n
 CC (where n is 1-15) of L-lysine amino acid residues of alpha and
 CC epsilon amide linkage, where (n+1)/2 of the alpha amino groups
 CC and/or (n+1)/2 of the epsilon amino groups are substituted with

CC polypeptides consisting of one or more plasmodial B epitopes
 CC covalently bound to one or more peptides with an amino acid sequence
 CC corresponding to that of a T epitope such as FNNFTVSWLRVPKVSASHLEA
 CC or QYKANSKFTGIVE
 CC The branched polypeptides can be used as immunogens for
 CC preparing genetically non-restricted antimalaria vaccines and for
 CC determining anti-plasmodium antibodies in blood, serum and blood-spot
 CC samples. Determination can be effected by ELISA.
 CC See also R14261 - R14263, R14265 and R15436.
 CC Sequence 77 AA;

Query Match 100.0%; Score 158; DB 3; Length 77;
 Best Local Similarity 100.0%; Pred. No. 4.34e-09;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 13 fnnftvswlrvpkvsashle 33
 QY 1 FNNFTVSWLRVPKVSASHLE 21
 |||||

RESULT 14
 ID R12471 standard; Protein; 452 AA.

AC R12471;
 DT 05-AUG-1991 (first entry)
 DE Tetanus toxin fragment C encoded by gene with increased G+C content.
 KW Terminator; vaccine.
 OS Synthetic.
 PN EP-430545-A.
 PD 05-JUN-1991.
 PF 27-NOV-1990; 312870.
 PR 28-NOV-1989; GB-026832.
 PR 17-MAR-1990; GB-006097.
 PA (WELL) WELLCOME FOUNDATION LTD.
 PI Makoff AJ, Romanos MA, Clare JT, Fairweather NF;
 DR WPI; 91-166115/23.
 DR N-PSDB; Q12121.
 PT DNA sequence encoding tetanus toxin fragment C - useful in the
 PT manufacture of vaccines for immunity to tetanus utilising yeast
 PT as host organism.
 PS Disclosure; Fig 2; 50pp; English.
 CC The (G+C) content of the synthetic gene is increased by 47% wrt the
 CC native sequence. This eliminates six "terminator" regions which
 CC were found to be present in (A+T) rich regions. The terminators
 CC (termination/endo-nucleolytic processing/polyadenylation sites)
 CC were previously responsible for incomplete transcription of the
 CC mRNA. The elimination of these elements (using codon degeneracy)
 CC provided for successful expression in yeast of the tetanus toxin
 CC fragment C.
 CC Sequence 452 AA;

Query Match 100.0%; Score 158; DB 2; Length 452;
 Best Local Similarity 100.0%; Pred. No. 4.34e-09;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 84 fnnftvswlrvpkvsashle 104
 QY 1 FNNFTVSWLRVPKVSASHLE 21
 |||||

RESULT 15
 ID P70345 standard; Protein; 573 AA.

AC P70345;
 DT 22-APR-1991 (first entry)
 DE Portion of B fragment and all of the C fragment of tetanus toxin.
 KW TI; vaccine.
 OS Clostridium tetani.
 PN EP-209281-A.
 PD 21-JAN-1987.
 PR 27-JUN-1986; 305029.
 PR 28-JUN-1985; GB-016442.
 PA (WELL) Wellcome Foundation Ltd.
 PI Fairweather NF;
 DR WPI; 87-015999/03.

DR N-PSDB; N70545.
 PT Cloned DNA sequence coding for tetanus toxin - or its fragments
 PT contg. epitope used to express antigens for vaccine production.
 PS Claim 4; Fig 1; 36pp; English.
 CC Gene product comprises a tetanus toxin fragment, which may be
 CC expressed in a transformed host, and used as an antigen in vaccine
 CC production, against the disease.
 CC Sequence 573 AA;

Query Match 100.0%; Score 158; DB 2; Length 573;
 Best Local Similarity 100.0%; Pred. No. 4.34e-09;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 205 fnnftvswlrvpkvsashle 225
 QY 1 FNNFTVSWLRVPKVSASHLE 21
 |||||

RESULT 16
 ID W48909 standard; Protein; 618 AA.

AC W48909;
 DT 23-SEP-1998 (first entry)
 DE SOD-1/TTC hybrid protein
 KW Chimeric; copper-zinc superoxide dismutase; SOD-1; TTC; SOD:Tet451;
 KW tetanus toxin fragment C; tetanus holotoxin; nerve cell; stroke;
 KW neurological disorder; oxidative stress; brain hypoxia-reperfusion;
 KW epilepsy; Parkinson's disease; Huntington's disease.
 OS Chimeric - Homo sapiens.
 OS Chimeric - Clostridium tetani.
 PR Key Location/Qualifiers
 FT Region 1..163
 FT /note= "SOD-1"
 FT Region 168..618
 FT /note= "TTC moiety"
 FT US5780024-A.
 PN 14-JUL-1998.
 PD 21-JUN-1996; 668381.
 PR 23-JUN-1995; US-000473.
 PR 21-JUN-1996; US-668381.
 PA (GHO) GEN HOSPITAL CORP.
 PA (UYMA-) UNIV MARYLAND BALTIMORE.
 PI Brown RH, Fishman PS, Francis JW, Hosler BA;
 DR WPI; 98-412999/35.
 DR N-PSDB; V32580.
 PT New hybrid protein of superoxide dismutase and tetanus toxin
 PT fragment C - having increased uptake by neurons and retention of
 PT enzymatic activity in these cells, for treating neurological
 PT diseases associated with oxidative stress
 PT Claim 7; Columns 23-26; 23pp; English.

CC The present sequence represents an enzymatically active human
 CC copper-zinc superoxide dismutase (SOD-1) fused at its carboxyl
 CC terminus with the tetanus toxin fragment C (TTC) moiety. The TTC
 CC moiety constitutes amino acid residues 865-1315 of the tetanus
 CC holotoxin. The hybrid protein, referred as SOD:Tet451, is claimed
 CC to have the following properties: (a) it exhibits Cu/Zn SOD enzymatic
 CC activity; (b) the TTC moiety selectively binds to nerve cells and
 CC allows uptake of the hybrid protein into these cells; and (c) it
 CC retains substantial SOD enzymatic activity following cellular uptake.
 CC SOD:Tet451 is claimed to be useful for treating neurological disorders
 CC associated with oxidative stress, e.g. stroke, brain hypoxia-reperfusion,
 CC epilepsy, Parkinson's and Huntington's diseases.

Query Match 100.0%; Score 158; DB 33; Length 618;
 Best Local Similarity 100.0%; Pred. No. 4.34e-09;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 250 fnnftvswlrvpkvsashle 270
 QY 1 FNNFTVSWLRVPKVSASHLE 21
 |||||

RESULT 17

ID R82574 standard; peptide; 22 AA.
 AC R82574;
 DT 13-JUN-1996 (first entry)
 DE Tetanus toxin helper T cell epitope, TT2.
 KW IgE; CH4; immunoglobulin; epsilon; immunogen; helper T cell; epitope;
 KW vaccine; allergy; antibody; constant heavy chain.
 OS Clostridium tetani.
 PN W09526365-A1.
 PD 05-OCT-1995.
 PF 24-MAR-1995; U03741.
 PR 28-MAR-1994; US-218461.
 PR 25-OCT-1994; US-328912.
 PA (UNBI-) UNITED BIOMEDICAL INC.
 PI Wang CY;
 DR WPI; 95-351297/45.
 PT Synthetic peptide-based immunogen contg. IgE CH4 peptide and helper
 PT T cell epitope - useful for eliciting antibody prodn. for allergy
 PT treatment
 PS Claim 3; Page 59; 87pp; English.
 CC R82571-91 are helper T cell epitopes which can be used in the
 CC preparation of a peptide immunogen that is useful in vaccines for
 CC treating allergic reactions. In the immunogen an IgE CH4 peptide
 CC is attached C-terminally to a series of amino acids including a
 CC helper T cell epitope. The immunogen may also opt. contain a fatty
 CC acid or fatty acid derivative, an invasin domain or alpha-NH2. The
 CC immunogen produces high titres of antibodies to the effector site
 CC in human IgE heavy chain (the CH4 domain peptide) which inhibit mast
 CC cell activation and reduce allergen-induced IgE prodn. The immunogens
 CC can be used in either a radially branching multimeric form or a
 CC linearly arranged monomeric form.
 SQ Sequence 22 AA;
 Query Match 96.2%; Score 152; DB 16; Length 22;
 Best Local Similarity 100.0%; Pred. No. 2.09e-08;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 3 fnnftvfwlrvpkvsashl 22
 QY 1 FNNFTVSEWLRVFKVSASHL 20
 RESULT 18
 ID R62693 standard; peptide; 22 AA.
 AC R62693;
 DT 10-SEP-1995 (first entry)
 DE Helper T cell epitope for use in universal immune stimulator.
 KW Helper T cell epitope; universal immune stimulator; invasin; hapten;
 KW vaccine; tetanus toxin.
 OS Clostridium tetani.
 PN W09425060-A.
 PD 10-NOV-1994.
 PF 28-APR-1994; U04832.
 PR 27-APR-1993; US-057166.
 PR 14-APR-1994; US-229275.
 PA (LADD/) LADD A E.
 PA (WANG/) WANG C Y.
 PA (ZAMB/) ZAMB I.
 PI Ladd AE, Wang CY, Zamb T;
 DR WPI; 94-357910/44.
 PT Immunogenic interleukinising hormone releasing hormone peptide(s) -
 PT that suppress LHRH activity in males and females
 PS Claim 7; Page 25; 213pp; English.
 CC Synthetic immunogenic peptides are provided in which a universal immune
 CC stimulator is linked to a peptide or protein hapten containing B cell
 CC and/or cytotoxic T lymphocyte epitopes, giving a product which causes
 CC potent immune responses to the coupled peptide or protein. The
 CC stimulator consists of (A) a promiscuous helper T cell epitope (Th)
 CC which elicits an immune response to the coupled peptide in members of
 CC a heterogeneous population expressing diverse HLA phenotypes, and (B)
 CC an adjuvant peptide sequence from the invasin protein of Yersinia.
 CC Spacer amino acid sequences (e.g. Gly-Gly) can be provided between the
 CC invasin and Th domains and between the immune stimulator and hapten
 CC components. When the hapten is LHRH, then optionally the invasin domain

CC can be omitted from the immune stimulator component.
 CC The present sequence represents a tetanus toxin helper T cell epitope
 CC which can be used as Th in the immune stimulator.
 SQ Sequence 22 AA;
 Query Match 96.2%; Score 152; DB 13; Length 22;
 Best Local Similarity 100.0%; Pred. No. 2.09e-08;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 3 fnnftvfwlrvpkvsashl 22
 QY 1 FNNFTVSEWLRVFKVSASHL 20
 RESULT 19
 ID W05600 standard; peptide; 22 AA.
 AC W05600;
 DT 10-DEC-1996 (first entry)
 DE Tetanus toxin helper T cell epitope #2.
 KW immunoglobulin; IgE; membrane protein; human; epsilon chain; hepatitis B;
 KW membrane anchoring domain; helper T cell; surface antigen; core antigen;
 KW pertussis toxin; tetanus toxin; measles virus F protein; immunotherapy;
 KW Chlamydia trachomatis major outer membrane protein; immunogen; vaccine;
 KW diphtheria toxin; plasmidium falciparum; circumsporozoite; E. coli Trat;
 KW schistosoma mansoni; triose phosphate isomerase; allergenic reaction;
 KW allergic rhinitis; food allergy; anaphylaxis; virally-induced asthma;
 KW antihistamine; decongestant; beta-2 agonist; immunosuppression;
 KW corticosteroid.
 OS Synthetic.
 PN W09612740-A1.
 PD 02-MAY-1996.
 PF 25-OCT-1995; U13841.
 PR 25-OCT-1994; US-328519.
 PA (UNBI-) UNITED BIOMEDICAL INC.
 PI Walfield AM, Wang CY;
 DR WPI; 96-230555/23.
 PT Peptide immunogen useful in treatment of allergy - comprises
 PT membrane-bound IgE epsilon-chain peptide synthesised linearly in
 PT tandem with T helper epitope peptide
 PS Claim 2; Page 18; 53pp; English.
 CC W05957-W05616 represent helper T cell epitopes used in the peptide
 CC immunogens of the invention. The peptides of the invention contain one of
 CC these sequences, and a membrane-bound immunoglobulin E (IgE) fragment
 CC (see W05595 and W05596). The peptide immunogens of the invention can be
 CC used in vaccines for the immunotherapeutic treatment of allergic
 CC reactions, including allergic rhinitis, food allergies, anaphylaxis, or
 CC virally-induced asthma. The immunogens overcome the short effective
 CC period of antihistamines, decongestants, and beta-2 agonists, while
 CC preventing the broad immunosuppression of corticosteroids. The peptides
 CC do not have the potential side effects of restlessness or sedation
 CC (associated with antihistamines), associated increased morbidity in
 CC asthmatics (as seen with beta-2 agonists) and adverse hormonal activities
 CC (observed in corticosteroid users).
 SQ Sequence 22 AA;
 Query Match 96.2%; Score 152; DB 19; Length 22;
 Best Local Similarity 100.0%; Pred. No. 2.09e-08;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 3 fnnftvfwlrvpkvsashl 22
 QY 1 FNNFTVSEWLRVFKVSASHL 20
 RESULT 20
 ID R82597 standard; peptide; 32 AA.
 AC R82597;
 DT 13-JUN-1996 (first entry)
 DE IgE CH4 region contg. peptide immunogen for treating allergies.
 KW IgE; CH4; immunoglobulin; epsilon; immunogen; helper T cell; epitope;
 KW vaccine; allergy; antibody; constant heavy chain.
 OS Synthetic.

Search completed: Tue Aug 17 16:10:56 1999
Job time : 29 secs.

```

PN WO9526365-A1.
PD 05-OCT-1995.
PF 24-MAR-1995; U03741.
PR 28-MAR-1994; US-218461.
PR 25-OCT-1994; US-328912.
PA (UNBI-) UNITED BIOMEDICAL INC.
PI Wang CY;
DR WPI; 95-351297/45.
PT Synthetic peptide-based immunogen contg. IgE CH4 peptide and helper
PT T cell epitope - useful for eliciting antibody prodn. for allergy
PT treatment
PS Claim 5; Page 63; 87pp; English.
CC R82592-R82600 and R83560-R83581 are peptide immunogens that are
CC useful in vaccines for treating allergic reactions. In the immunogens,
CC an IgE CH4 peptide is attached C-terminally to a series of amino acids
CC including a helper T cell epitope. The immunogen may also opt. contain
CC a fatty acid or fatty acid derivative, an invasin domain or alpha-NH2.
CC The immunogen produces high titres of antibodies to the effector site
CC in human IgE heavy chain (the CH4 domain peptide) which inhibit mast
CC cell activation and reduce allergen-induced IgE prodn. The immunogens
CC may be used in either a radially branching multimeric form or a
CC linearly arranged monomeric form.
SQ Sequence 32 AA;

Query Match 96.2%; Score 152; DB 16; Length 32;
Best Local Similarity 100.0%; Pred. No. 2.09e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

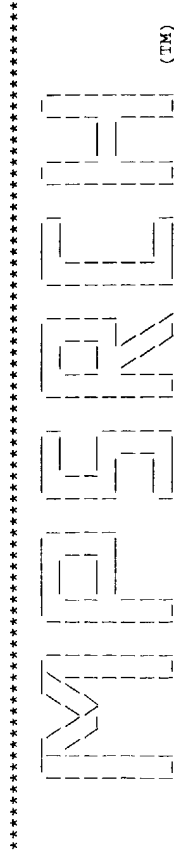
Db 3 fnftvswlrpkvsashl 22
   |||||
QY 1 FNNFTVSWLRPKVSASHL 20

RESULT 21
ID R83562 standard; peptide; 34 AA.
AC R83562;
DT 13-JUN-1996 (first entry)
DE IgE CH4 region contg. peptide immunogen for treating allergies.
KW IgE; CH4; immunoglobulin; epsilon; immunogen; helper T cell; epitope;
KW vaccine; allergy; antibody; constant heavy chain.
OS Synthetic.
PN WO9526365-A1.
PD 05-OCT-1995.
PF 24-MAR-1995; U03741.
PR 28-MAR-1994; US-218461.
PR 25-OCT-1994; US-328912.
PA (UNBI-) UNITED BIOMEDICAL INC.
PI Wang CY;
DR WPI; 95-351297/45.
PT Synthetic peptide-based immunogen contg. IgE CH4 peptide and helper
PT T cell epitope - useful for eliciting antibody prodn. for allergy
PT treatment
PS Claim 5; Page 69; 87pp; English.
CC R82592-R82600 and R83560-R83581 are peptide immunogens that are
CC useful in vaccines for treating allergic reactions. In the immunogens,
CC an IgE CH4 peptide is attached C-terminally to a series of amino acids
CC including a helper T cell epitope. The immunogen may also opt. contain
CC a fatty acid or fatty acid derivative, an invasin domain or alpha-NH2.
CC The immunogen produces high titres of antibodies to the effector site
CC in human IgE heavy chain (the CH4 domain peptide) which inhibit mast
CC cell activation and reduce allergen-induced IgE prodn. The immunogens
CC may be used in either a radially branching multimeric form or a
CC linearly arranged monomeric form.
SQ Sequence 34 AA;

Query Match 96.2%; Score 152; DB 16; Length 34;
Best Local Similarity 100.0%; Pred. No. 2.09e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3 fnftvswlrpkvsashl 22
   |||||
QY 1 FNNFTVSWLRPKVSASHL 20

```



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Tue Aug 17 15:47:39 1999; MasPar time 4.52 Seconds
Tabular output not generated. 132.990 Million cell updates/sec

Title: >US-09-049-847-1
Description: (1-15) from US09049847.ppep
Perfect Score: 104
Sequence: 1 QVIKANSKFIGITEL 15

Scoring table: PAM 150
Gap 15

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 26.532; Variance 37.227; scale 0.713

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	104	100.0	1315	1	BTCCLN	tentoxylisin (EC 3.4. 6.52e-10
2	63	60.6	213	1	KTYMC	adenylate kinase (EC 5.01e-01
3	63	60.6	899	2	G36812	hypothetical protein 7.01e-01
4	62	59.6	284	2	S77138	hypothetical protein 7.75e-01
5	62	59.6	598	2	F69792	hypothetical protein 7.75e-01
6	61	58.7	194	2	G64026	[acyl-carrier-protein 1.19e+00
7	60	57.7	568	2	S05532	gamma-glutamyltransfe 1.83e+00
8	59	56.7	66	2	S31029	gene 84 protein - Myc 2.79e+00
9	59	56.7	123	2	G48677	Ig heavy chain V-D-J 2.79e+00
10	59	56.7	287	2	F70361	tRNA pseudouridine 55 2.79e+00
11	58	55.8	326	2	B71808	type II restriction e 4.24e+00
12	58	55.8	357	2	C69908	delta-endotoxin homol 4.24e+00
13	58	55.8	423	2	F64590	type IIS restriction 4.24e+00
14	57	54.8	269	2	S73999	hypothetical protein 6.40e+00
15	57	54.8	322	2	B64382	formylmethanofuran de 6.40e+00
16	57	54.8	400	2	A45545	major merozoite surfa 6.40e+00
17	57	54.8	1333	2	S38635	blastopia polypeptide 6.40e+00
18	57	54.8	1531	1	SA20K1	major merozoite surfa 6.40e+00
19	57	54.8	1639	2	S05603	major merozoite surfa 6.40e+00
20	57	54.8	1840	2	A24594	probable major surfac 6.40e+00
21	56	53.8	227	2	S72285	ribosomal protein S2 9.63e+00
22	56	53.8	316	2	D69692	riboflavin kinase / F 9.63e+00
23	56	53.8	381	2	F71196	hypothetical protein 9.63e+00

ALIGNMENTS

RESULT	ENTRY	BTCLTN	#type complete
24	56	601	2
25	53.8	85	2
26	52.9	102	2
27	52.9	119	2
28	52.9	119	2
29	52.9	119	2
30	52.9	123	2
31	52.9	135	2
32	52.9	140	2
33	52.9	165	2
34	52.9	290	2
35	52.9	382	2
36	52.9	461	2
37	52.9	757	2
38	52.9	1146	2
39	52.9	1182	2
40	54	108	2
41	54	499	2
42	54	501	2
43	54	644	2
44	53	1060	2
45	53	7962	2
ALTERNATE_NAMES	tentoxylisin (EC 3.4.24.68) precursor - Clostridium tetani		
ORGANISM	#formal_name Clostridium tetani		
DATE	31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 26-Feb-1999		
ACCESSIONS	A25689; A25757; A25194; B25194; A60759; S69348; S09364		
REFERENCE	A25689		
#authors	Eisel, U.; Jarausch, W.; Goretzki, K.; Henschel, A.; Engels, J.; Weller, U.; Hudel, M.; Habermann, E.; Niemann, H.		
#journal	EMBO J. (1986) 5:2495-2502		
#title	Tetanus toxin: primary structure, expression in E. coli, and homology with botulinum toxins.		
#cross-references	MUID:87053814		
#accession	A25689		
#molecule_type	DNA		
#residues	1-1315	#label EIS	
#cross-references	GB:X04436; NID:940769; PID:g40770		
REFERENCE	A25757		
#authors	Fairweather, N.F.; Lyness, V.A.		
#journal	Nucleic Acids Res. (1986) 14:7809-7812		
#title	The complete nucleotide sequence of tetanus toxin.		
#cross-references	MUID:87040747		
#accession	A25757		
#molecule_type	DNA		
#residues	1-1315	#label FAI	
#cross-references	GB:X06214; NID:940773; PID:g40774		
#experimental_source	strain CN3911		
REFERENCE	A25194		
#authors	Fairweather, N.F.; Lyness, V.A.; Pickard, D.J.; Allen, G.; Thomson, R.O.		
#journal	J. Bacteriol. (1986) 165:21-27		
#title	Cloning, nucleotide sequencing, and expression of tetanus toxin fragment C in Escherichia coli.		
#cross-references	MUID:86085672		
#accession	A25194		
#molecule_type	DNA		
#residues	743-1315	#label FA2	
#cross-references	GB:M12739; NID:944920; PID:g144921		
#accession	B25194		
#molecule_type	protein		
#residues	865-894	#label FA3	
REFERENCE	A60759		
#authors	Matsuda, M.; Lei, D.L.; Sugimoto, N.; Ozutsumi, K.; Okabe, T.		
#journal	Infect. Immun. (1989) 57:3588-3593		
#title	Isolation, purification, and characterization of fragment B,		

```

the NH-2-terminal half of the heavy chain of tetanus toxin.
#cross-references MUID:90035436
#accession A60759
#molecule_type protein
#residues 461-475 #label MAT
REFERENCE JS0098
#authors Denotz, S.; Lanzavechia, L.; Eisel, U.; Niemann, H.;
#journal J. Immunol. (1989); 142:394-402
#title Delineation of several DR-restricted tetanus toxin T cell
#molecule_type DNA
#cross-references MUID:89093918
#contents annotation; epitope region
#journal S27125
#authors Schiavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de
#journal Nature (1992) 359:832-835
#title Tetanus and botulinum-B neurotoxins block neurotransmitter
#molecule_type DNA
#cross-references MUID:93063293
#contents annotation
#journal S69348
#authors de Filippis, V.; Vangelista, L.; Schiavo, G.; Tonello, F.;
#journal Eur. J. Biochem. (1995) 229:61-69
#title Structural studies on the zinc-endopeptidase light chain of
#molecule_type protein
#cross-references MUID:95262688
#accession S69348
#residues 2-31 #label DEF
COMMENT The source of this protein was an extrachromosomal plasmid.
COMMENT The precursor is cleaved by endogenous proteinase activity to form
light (fragment A) and heavy (fragment B.C) chains that are
covalently linked by an interchain disulfide bond (the individual
chains are not toxic when separated). The amino end of the heavy
chain (fragment B) can be separated from the carboxyl end
(fragment C) by papain.
COMMENT Fragment B forms ion channels in a lipid bilayer. Fragment C binds
to gangliosides and may target the toxin to the motor end plate.
COMMENT Fragment A is a zinc-dependent endopeptidase.
COMMENT This potent neurotoxin binds to peripheral neuronal synapses, is
internalized, and moves by retrograde transport up the axon into
the spinal cord where it can move between postsynaptic and
presynaptic neurons. It inhibits neurotransmitter release by
proteolytic cleavage of synaptobrevin (vesicle-associated
membrane protein 2).
FUNCTION
#description blocks neuroexocytosis via hydrolysis of a Gln-Phe peptide
bond in synaptobrevin 2
CLASSIFICATION #superfamily tetanus toxin
KEYWORDS hydrolase; metalloproteinase; neurotoxin; transmembrane
protein; zinc
FEATURE
2-457 #product tentoxylisin light chain (fragment A) #status
#predicted #label IIN
461-1315 #product tentoxylisin heavy chain (fragment B.C) #status
#experimental #label TTH
461-864 #domain channel forming (fragment B) #status predicted
#label IXH
865-1315 #domain ganglioside binding (fragment C) #status
#predicted #label IXC
233,237 #binding_site zinc (His) #status predicted\
234 #active_site Glu #status predicted
SUMMARY #length 1315 #molecular-weight 150681 #checksum 4853
Query Match 100.0%; Score 104; DB 1; Length 1315;
Best Local Similarity 100.0%; Pred. No. 6.52e-10;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 830 QYKANSKFTGITEL 844
QY I QYKANSKFTGITEL 15

```

```

RESULT 2
ENTRY KIYMC #type complete
TITLE adenylate kinase (EC 2.7.4.3) - Mycoplasma capricolum (SGC3)
ORGANISM #formal_name Mycoplasma capricolum
DATE 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change
05-Sep-1997
ACCESSIONS S02851
REFERENCE S02830
#authors Ohkubo, S.; Muto, A.; Kawachi, Y.; Yamao, F.; Osawa, S.
#journal Mol. Gen. Genet. (1987) 210:314-322
#title The ribosomal protein gene cluster of Mycoplasma capricolum.
#cross-references MUID:88142549
#accession S02851
#molecule_type DNA
#residues 1-213 #label OHK
#cross-references EMBL:X06414; NID:g44207; PID:g44229
GENETICS
#gene adk
#genetic_code SGC3
FUNCTION #description catalyzes the reversible phosphorylation of adenine
monophosphate with adenosine triphosphate to form two
adenosine diphosphates in the presence of magnesium
#superfamily adenylate kinase
KEYWORDS ATP; P-loop; phosphotransferase
FEATURE
7-14 #region nucleotide-binding motif A (P-loop) #status
atypical\
78-83 #region nucleotide-binding motif B #status atypical\
30,82 #active_site Ser, Asp #status predicted
SUMMARY #length 213 #molecular-weight 24616 #checksum 5992
Query Match 60.68; Score 63; DB 1; Length 213;
Best Local Similarity 66.7%; Pred. No. 5.01e-01;
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 182 DYKTNKSKFTETI 193
QY 1 QYKANSKFTGI 12
RESULT 3
ENTRY G36812 #type complete
TITLE hypothetical protein ORF63 - saimirine herpesvirus 1 (strain
11)
ORGANISM #formal_name saimirine herpesvirus 1
#note host Saimiri sciureus (common squirrel monkey)
DATE 16-Oct-1992 #sequence_revision 16-Oct-1992 #text_change
09-Sep-1997
ACCESSIONS G36812
REFERENCE A38806
#authors Albrecht, J.
#submission submitted to the EMBL Data Library, January 1992
#description Primary structure of the herpesvirus saimirine genome.
#accession G36812
#molecule_type DNA
#residues 1-899 #label ALP
#cross-references GB:X64346; NID:g60320; PID:g60384
REFERENCE A37309
#authors Albrecht, J.C.; Nicholas, J.; Biller, D.; Cameron, K.R.;
Biesinger, B.; Newman, C.; Wittmann, S.; Craxton, M.A.;
Coleman, H.; Fleckenstein, B.; Honess, R.W.
#journal J. Virol. (1992) 66:5047-5058
#title Primary structure of the herpesvirus saimirine genome.
#cross-references MUID:92333688
#contents annotation; protein-coding frames
#note neither protein nor nucleotide sequence is given
GENETICS
#gene
SUMMARY #length 899 #molecular-weight 103350 #checksum 542

```

```

Query Match      60.6%; Score 63; DB 2; Length 899;
Best Local Similarity 50.0%; Pred. No. 5.01e-01;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 124 QYTSNATFTGLSE 137
QY 1 QYIKANSKFIGITE 14
      ||| ||| ||| |||
      ||| ||| ||| |||

RESULT 4
ENTRY S77138 #type complete
TITLE hypothetical protein sir1882 - Synecocystis sp. (strain PCC 6803)
ORGANISM #formal_name Synecocystis sp.
DATE 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 21-Aug-1998
ACCESSIONS S77138
REFERENCE S74322
#authors Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugiyama, M.; Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Tabata, S.
#journal DNA Res. (1996) 3:109-136
#title Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis sp. PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.
#cross-references MUID:97061201
#accession S77138
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-284 #label KAN
#cross-references EMBL:D90908; GB:AB001339; NID:g1652725; PID:d1018429; PID:g1652777
#note the nucleotide sequence was submitted to the EMBL Data Library, June 1996

GENETICS
#start_codon GTG
CLASSIFICATION #superfamily conserved hypothetical protein H10963
SUMMARY #length 284 #molecular-weight 31380 #checksum 3965

Query Match      59.6%; Score 62; DB 2; Length 284;
Best Local Similarity 33.3%; Pred. No. 7.75e-01;
Matches 5; Conservative 8; Mismatches 2; Indels 0; Gaps 0;

Db 248 HYLPRPKFAGLDL 262
QY 1 QYIKANSKFIGITE 15
      ||| ||| ||| |||
      ||| ||| ||| |||

RESULT 5
ENTRY F69792 #type complete
TITLE hypothetical protein yeeB - Bacillus subtilis
ORGANISM #formal_name Bacillus subtilis
DATE 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 24-Sep-1998
ACCESSIONS F69792
REFERENCE A69580
#authors Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.; Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans, A.; Braun, M.; Brigneau, S.C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.; Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Galizzi, A.; Galleron, N.; Ghim, S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.; Guiseppe, G.; Guy, B.J.; Haga, K.; Haiech, J.; Harwood,

```

```

C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, B.; Joris, B.; Karamata, D.; Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetle, D.; Porwollik, S.; Prescott, G.; A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Repoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.; Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenegger, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
#journal Nature (1997) 390:249-256
#title The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
#cross-references MUID:98044033
#accession F69792
#status preliminary; nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-598 #label KUN
#cross-references GB:Z99107; GB:AL009126; NID:g2632866; PID:el182657; PID:g2632991
#experimental_source strain 168

GENETICS
#gene yeeB
SUMMARY #length 598 #molecular-weight 67963 #checksum 7652

Query Match      59.6%; Score 62; DB 2; Length 598;
Best Local Similarity 40.0%; Pred. No. 7.75e-01;
Matches 6; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Db 531 EFIRMDKFNIEDL 545
QY 1 QYIKANSKFIGITE 15
      ||| ||| ||| |||
      ||| ||| ||| |||

RESULT 6
ENTRY G64026 #type complete
TITLE [acyl-carrier-protein] phosphodiesterase (EC 3.1.4.14) H11366
ALTERNATE_NAMES - Haemophilus influenzae
ORGANISM conserved hypothetical protein H11366
DATE 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 21-Nov-1998
ACCESSIONS G64026
REFERENCE A64000
#authors Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kervage, A.R.; Bult, C.J.; Tomb, J.F.; Dougherty, B.A.; Merrick, J.M.; McKenney, K.; Sutton, G.; Fitzhugh, W.; Fields, C.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.F.; Phillips, C.A.; Spriggs, T.; Hedblom, E.; Cotton, M.D.; Utterback, T.R.; Hanna, M.C.; Nguyen, D.T.; Saudak, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M.; Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, J.C.
#journal Science (1995) 269:496-512
#title Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
#cross-references MUID:95350630
#accession G64026

```

```

##status      nucleic acid sequence not shown; translation not shown
##molecule_type DNA
##residues    1-194 #label TIGR
##cross-references GB:U32816; GB:L42023; NID:g1574193; PID:g1574198;
##experimental_source strain Rd KW20
FUNCTION
#description   catalyzes hydrolysis of the phosphopantetheine residue from
               holo-acyl-carrier-protein
CLASSIFICATION #superfamily acyl carrier protein phosphodiesterase
KEYWORDS       phosphoric diester hydrolase
SUMMARY        #length 194 #molecular-weight 21208 #checksum 8859
Query Match    58.7%; Score 61; DB 2; Length 194;
Best Local Similarity 53.3%; Pred. No. 1.19e+00;
Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
Db 147 QYMKSLGFIQTDV 161
|:|:| | | | | |
QY 1 QYIKANSKFIGITEL 15

RESULT 7
ENTRY   S05532 #type complete
TITLE   gamma-glutamyltransferase (EC 2.3.2.2) - pig
ALTERNATE_NAMES gamma-glutamyl transpeptidase
ORGANISM #formal_name Sus scrofa domestica #common_name domestic pig
DATE     30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change
        29-Jan-1999
ACCESSIONS S05532
REFERENCE   S05532
#authors   Papadrikopoulou, A.; Frey, A.; Gassen, H.G.
#journal   Eur. J. Biochem. (1989) 183:693-698
#title     Cloning and expression of gamma-glutamyl transpeptidase from
           isolated porcine brain capillaries.
#cross-references MUID:89377838
#accession S05532
##status   not compared with conceptual translation
##molecule_type DNA
##residues 1-568 #label PAP
##cross-references GB:Z46922; NID:g600818; PID:g600819
CLASSIFICATION #superfamily gamma-glutamyltransferase
KEYWORDS       aminocyclotransferase; glycoprotein; heterodimer; membrane
               protein
FEATURE
1-379 #product gamma-glutamyltransferase heavy chain #status
380-568 #predicted #label HCH\
        #product gamma-glutamyltransferase light chain #status
        #predicted #label LCH
SUMMARY        #length 568 #molecular-weight 61315 #checksum 3041
Query Match    57.7%; Score 60; DB 2; Length 568;
Best Local Similarity 42.9%; Pred. No. 1.83e+00;
Matches 6; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
Db 531 HYIQDASTFQVVO 544
|:| | | | | |
QY 1 QYIKANSKFIGITE 14

RESULT 8
ENTRY   S31029 #type complete
TITLE   gene 84 protein - Mycobacterium phage L5
ORGANISM #formal_name Mycobacterium phage L5
DATE     30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change
        09-Sep-1997
ACCESSIONS S31029
REFERENCE   S30949
#authors   Donnelly-Wu, M.K.; Jacobs Jr., W.R.; Hatfull, G.F.
#journal   Mol. Microbiol. (1993) 7:407-417
#title     Superinfection immunity of mycobacteriophage L5: applications
           for genetic transformation of mycobacteria.
#accession S31029

```

```

##status      nucleic acid sequence not shown; translation not shown
##molecule_type DNA
##residues    1-66 #label DON
##cross-references EMBL:Z18946; NID:g15859; PID:e59702; PID:g579152
##note        the nucleotide sequence was submitted to the EMBL Data
               Library, December 1992
GENETICS
#gene         84
#start_codon GTG
SUMMARY        #length 66 #molecular-weight 7424 #checksum 8203
Query Match    56.7%; Score 59; DB 2; Length 66;
Best Local Similarity 70.0%; Pred. No. 2.79e+00;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 50 YIKRNGKFGV 59
| | | | | | |
QY 2 YIKANSKFIG 11

RESULT 9
ENTRY   G48677 #type fragment
TITLE   Ig heavy chain V-D-J region (419.1) - mouse (fragment)
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE     19-May-1994 #sequence_revision 19-May-1994 #text_change
        17-Mar-1999
ACCESSIONS G48677
REFERENCE   A48677
#authors   Tassignon, J.; Brait, M.; Jamila, I.; Urbain, J.; Gottlieb,
           P.; Brown, A.; Hasemann, C.A.; Capra, J.D.; Meek, K.
#journal   Proc. Natl. Acad. Sci. U.S.A. (1993) 90:9508-9512
#title     Molecular characterization of monoclonal CRI-A-positive
           anti-arsonate antibodies derived from idiotypic-negative
           mice bearing a light chain polymorphism.
#cross-references MUID:94022404
#accession G48677
##status   preliminary; not compared with conceptual translation
##molecule_type mRNA
##residues 1-123 #label TAS
CLASSIFICATION #superfamily immunoglobulin V region; immunoglobulin homology
               heterotetramer; immunoglobulin
KEYWORDS       15-98
FEATURE        #domain immunoglobulin homology #label IMM
SUMMARY        #length 123 #checksum 1208
Query Match    56.7%; Score 59; DB 2; Length 123;
Best Local Similarity 60.0%; Pred. No. 2.79e+00;
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
Db 56 DYIKYNEKFGITTL 70
| | | | | | |
QY 1 QYIKANSKFIGITEL 15

RESULT 10
ENTRY   F70361 #type complete
TITLE   tRNA pseudouridine 55 synthase - Aquifex aeolicus
ORGANISM #formal_name Aquifex aeolicus
DATE     08-May-1998 #sequence_revision 08-May-1998 #text_change
        21-Aug-1998
ACCESSIONS F70361
REFERENCE   A70300
#authors   Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.;
           Lenox, A.L.; Graham, D.E.; Overbeek, R.; Snead, M.A.;
           Keller, M.; AuJay, M.; Huber, R.; Feldman, R.A.; Short,
           J.M.; Olson, G.J.; Swanson, R.V.
#journal   Nature (1998) 392:353-358
#title     The complete genome of the hyperthermophilic bacterium
           Aquifex aeolicus.
#cross-references MUID:98195666
#accession F70361
##status   preliminary; nucleic acid sequence not shown;
           translation not shown

```

```

##molecule_type DNA
##residues 1-287 ##label AOF
##cross-references GB:AE000703; NID:g2983287; PID:g2983293; GB:AE000657
##experimental_source strain VF5

GENETICS
#gene
#superfamily Escherichia coli protein p35
CLASSIFICATION
#length 287 #molecular-weight 32259 #checksum 7335
SUMMARY
Query Match 56.7%; Score 59; DB 2; Length 287;
Best Local Similarity 80.0%; Pred. No. 2.79e+00;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 266 DSKFIGIGTEL 275
:|||||
6 NSKFIGITEL 15

RESULT 11
ENTRY
TITLE type II restriction enzyme - Helicobacter pylori (strain J99)
ORGANISM #formal_name Helicobacter pylori
#variety strain J99
DATE 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change
12-Feb-1999

ACCESSIONS B71808
REFERENCE A71800
#authors Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.;
Doig, P.C.; Smith, D.R.; Noonan, B.; Guild, B.C.; deJonge,
B.L.; Carmel, G.; Tummino, P.J.; Caruso, A.;
Uria-Nickelsen, M.; Mills, D.M.; Ives, C.; Gibson, R.;
Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis,
G.F.; Irust, T.J.

#journal
#title Nature (1999) 397:176-180
#cross-references MUID:99120557
#accession B71808
##status preliminary
##molecule_type DNA
##residues 1-326 ##label ARN
##cross-references GB:AB001566; GB:AE001439; NID:g4156051; PID:g4156064
##experimental_source strain J99

GENETICS
#gene
#length 326 #molecular-weight 39034 #checksum 2234
SUMMARY
Query Match 55.8%; Score 58; DB 2; Length 326;
Best Local Similarity 50.0%; Pred. No. 4.24e+00;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 43 YIOSNKKYISLTPL 56
:|||||
2 YIKANSKFIGITEL 15

RESULT 12
ENTRY
TITLE delta-endotoxin homolog yokG - Bacillus subtilis
ORGANISM #formal_name Bacillus subtilis
DATE 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
24-Sep-1998

ACCESSIONS C69908
REFERENCE A69580
#authors Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
Bolojin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
Choi, S.K.; Codani, J.F.; Connerton, I.F.; Cummings, N.J.;
Daniel, R.A.; Denizot, F.; Devigne, K.M.; Duesterhoeft, A.;
Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.;
Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,

```

```

M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
Guiseppi, G.; Guy, B.J.; Haga, K.; Haiech, J.; Harwood,
C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
Y.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.;
Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
M.; Osawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,
V.; Pohl, T.M.; Portetelie, D.; Porwollik, S.; Prescott,
A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.;
Schleich, S.; Schroeter, R.; Scoffone, E.; Sekiguchi, J.;
Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,
B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
Takamaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenegeer, T.;
Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
Yoshikawa, H.; Danchin, A.

#journal
#title Nature (1997) 390:249-256
#cross-references MUID:98044033
#accession C69908
##status preliminary; nucleic acid sequence not shown;
translation not shown

##molecule_type DNA
##residues 1-357 ##label KUN
##cross-references GB:Z99115; NID:g2634478; PID:eil83607;
PID:g2634580
##experimental_source strain 168

GENETICS
#gene
#length 357 #molecular-weight 40742 #checksum 1804
SUMMARY
Query Match 55.8%; Score 58; DB 2; Length 357;
Best Local Similarity 35.7%; Pred. No. 4.24e+00;
Matches 5; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

Db 165 FLOGGNFIOVTOL 178
:|||||
2 YIKANSKFIGITEL 15

RESULT 13
ENTRY
TITLE type IIS restriction enzyme R protein - Helicobacter pylori
ORGANISM #formal_name Helicobacter pylori
DATE 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change
10-Oct-1997

ACCESSIONS F64690
REFERENCE A64520
#authors Sutton, G.G.; Fleischmann, R.D.; Ketchum, K.A.; Klenk,
H.P.; Gill, S.; Dougherty, B.A.; Nelson, K.; Quackenbush,
J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus, B.;
Richardson, D.; Dodson, R.; Khalaf, H.G.; Glodek, A.;
McKenney, K.; Fitzgerald, L.M.; Lee, N.; Adams, M.D.;
Hickey, E.K.; Berg, D.E.; Gocayne, J.D.; Utterback, J.R.;
Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.;
Fujii, C.; Bowman, C.; Watthey, L.; Wallin, E.; Hayes,
W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,
C.M.; Venter, J.C.

#journal
#title Nature (1997) 388:539-547
The complete genome sequence of the gastric pathogen

```

```

#cross-references MUID:97394467
#accession F64690
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-423 #label TOM
#cross-references GB:AE000637; GB:AE000511; NID:g2314536; PID:g2314537;
TIGR:HP1366
SUMMARY #length 423 #molecular-weight 50047 #checksum 4262
Query Match 55.8%; Score 58; DB 2; Length 423;
Best Local Similarity 50.0%; Pred. No. 4.24e+00;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
Db 140 YIOSNKYISLTP 153
||:|:|:|:|
QY 2 YIKANSKFIGITEL 15

RESULT 14
ENTRY S73999 #type complete
TITLE hypothetical protein yaaC homolog VxpSPT7_orf269 - Mycoplasma
pneumoniae (ATCC 29342) (SGC3)
ALTERNATE_NAMES hypothetical protein VxpSPT7_orf269
ORGANISM #formal_name Mycoplasma pneumoniae
#variety ATCC 29342
DATE 27-Feb-1997 #sequence_revision 25-Apr-1997 #text_change
21-Aug-1998
ACCESSIONS S73999
REFERENCE S73327
#authors Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkel, E.; Li,
B.C.; Herrmann, R.
#journal Nucleic Acids Res. (1996) 24:4420-4449
#title Complete sequence analysis of the genome of the bacterium
Mycoplasma pneumoniae.
#cross-references MUID:97105885
#accession S73999
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-269 #label HIM
#cross-references EMBL:AE000062; GB:U00089; NID:g1674373; PID:g1674379
#note the nucleotide sequence was submitted to the EMBL Data
Library, November 1996
GENETICS
#genetic_code SGC3
CLASSIFICATION #superfamily conserved hypothetical protein HI0963
SUMMARY #length 269 #molecular-weight 30435 #checksum 2148
Query Match 54.8%; Score 57; DB 2; Length 269;
Best Local Similarity 40.0%; Pred. No. 6.40e+00;
Matches 5; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
Db 235 QFIRPQKXFGVOAL 249
||:|:|:|
QY 1 QYIKANSKFIGITEL 15

RESULT 15
ENTRY B64382 #type complete
TITLE formylmethanofuran dehydrogenase (tungsten) (EC 1.2.99.-)
ORGANISM subunit C related protein - Methanococcus jannaschii
#formal_name Methanococcus jannaschii
DATE 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change
10-Oct-1997
ACCESSIONS B64382
REFERENCE A64300
#authors Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann,
R.D.; Sutton, G.G.; Blake, J.A.; Fitzgerald, L.M.; Clayton,
R.A.; Gocayne, J.D.; Kerlavage, A.R.; Dougherty, B.A.;
Tomb, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.;
Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;

```

```

Scott, J.L.; Geoghagen, N.S.M.; Weidman, J.F.; Fuhrmann,
J.L.; Nguyen, D.; Utterback, T.R.; Kelley, J.M.; Peterson,
J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts,
K.M.; Hurst, M.A.; Kaine, B.P.; Borodovsky, M.; Klerk,
H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.
Science (1996) 273:1058-1073
#journal Complete genome sequence of the methanogenic archaeon,
Methanococcus jannaschii.
#title
#cross-references MUID:96337999
#accession B64382
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-322 #label BUL
#cross-references GB:U67513; GB:L77117; NID:g1591365; PID:g1591371;
TIGR:MU0658; PID:g1510742
GENETICS
#map_position FOR584512-585480
#start_codon TTG
KEYWORDS oxidoreductase
SUMMARY #length 322 #molecular-weight 36149 #checksum 2742
Query Match 54.8%; Score 57; DB 2; Length 322;
Best Local Similarity 61.5%; Pred. No. 6.40e+00;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
Db 153 IKGYRKFIITTEF 165
||:|:|:|:|
QY 3 IKANSKFIGITEL 15

```

Search completed: Tue Aug 17 15:48:10 1999
Job time : 31 secs.

(TM)

Result No.	Score	Query		ID	Description	Pred. No.	
		Match	Length				
1	104	100.0	1314	1	TETX_MYCCE	TETANUS TOXIN PRECURSOR	3.53e-12
2	63	60.6	213	1	KAD_WLOCA	ADENYLATE KINASE (EC 2	8.08e-02
3	63	60.6	899	1	V120_HSVSA	CAPSID ASSEMBLY PROTEI	8.08e-02
4	62	59.6	284	1	R1BF_SYN3A	RIBOFLAVIN KINASE (EC	1.34e-01
5	61	58.7	194	1	ACPD_HAEIN	ACYL CARRIER PROTEIN P	2.20e-01
6	60	57.7	568	1	GGL_PIG	GAMMA-GLUTAMYLTRANSEP	3.61e-01
7	59	56.7	66	1	V684_BPMLS	GENE 84 PROTEIN (GP84)	5.87e-01
8	59	56.7	287	1	TRUB_AQUAE	TRNA PSEUDOURIDINE SYN	5.87e-01
9	57	54.8	269	1	R1BF_MYCPN	PUTATIVE RIBOFLAVIN KI	1.53e+00
10	57	54.8	333	1	R1PB_XENIA	DNA POLYMERASE BETA (E	1.53e+00
11	57	54.8	1630	1	MSP1_PLAFK	MEROZOITE SURFACE PROT	1.53e+00
12	57	54.8	1639	1	MSP1_PLAFW	MEROZOITE SURFACE PROT	1.53e+00
13	56	53.8	316	1	R1BC_PACSU	RIBOFLAVIN KINASE (EC	2.44e+00
14	56	53.8	601	1	PPPF_LACLA	OLIGOENDOPEPTIDASE F (2.44e+00
15	56	53.8	887	1	ACOC_CAEEL	PROBABLE ACETITATE HYD	2.44e+00
16	55	52.9	85	1	Y668_METJA	HYPOTHETICAL PROTEIN M	3.88e+00
17	55	52.9	451	1	N1FN_RHOCA	NITROGENASE IRON-MOLIB	3.88e+00
18	55	52.9	757	1	R1S1_YEAST	R1S1 PROTEIN (SCS1 PRO	3.88e+00
19	55	52.9	1182	1	ABU2_HUMAN	TYROSINE-PROTEIN KINAS	6.13e+00
20	54	51.9	501	1	DLDH_PEA	DIHYDROLIPOAMIDE DEHYD	6.13e+00
21	54	51.9	528	1	MDUC_PSEPU	BENZOYLFORMATE DECARBO	6.13e+00
22	54	51.9	644	1	YH39_YEAS	HYPOTHETICAL ALDEHYDE-	6.13e+00
23	53	51.0	997	1	VMT2_IAZ11	MATRIX PROTEIN M2	9.61e+00

RP IDENTIFICATION AS ZINC-PROTEASE.
 RX MEDLINE; 93010948.
 RA SCHIAVO G., POULAIN B., ROSSETTO O., BENFENATI F., TAUC L.,
 RA MONTECUCCO C.;
 RT "Tetanus toxin is a zinc protein and its inhibition of
 RT neurotransmitter release and protease activity depend on zinc.";
 RL EMBO J. 11:3577-3583(1992).
 RN [7]
 RP IDENTIFICATION OF SUBSTRATE.
 RX MEDLINE; 93063293.
 RA SCHIAVO G., BENFENATI F., POULAIN B., ROSSETTO O., DE LAURETO P.P.,
 RA DASGUPTA B.R., MONTECUCCO C.;
 RT "Tetanus and botulinum-B neurotoxins block neurotransmitter release
 RT by proteolytic cleavage of synaptobrevin.";
 RL NATURE 359:832-835(1992).
 RN [8]
 RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 874-1314.
 RX MEDLINE; 97475217.
 RA UMLAND T.C., WINGERT L.M., SWAMINATHAN S., FUREY W.F., SCHMIDT J.J.,
 RA SAX M.;
 RT "Structure of the receptor binding fragment HC of tetanus
 RT neurotoxin.";
 RL NAT. STRUCT. BIOL. 4:788-792(1997).
 CC -!- FUNCTION: TETANUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
 CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
 CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
 CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
 CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
 CC ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 76-GLN-|-PHE-77
 CC BOND OF SYNAPTOSOMAL-2.
 CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF 76-GLN-|-PHE-77 BOND IN
 CC SYNAPTOSOMAL-2.
 CC -!- SUBUNIT: THE PRECURSOR POLYPEPTIDE IS SUBSEQUENTLY CLEAVED TO
 CC YIELD SUBCHAINS L AND H. THESE REMAIN LINKED BY A DISULFIDE BRIDGE
 CC AND ARE NON-TOXIC AFTER SEPARATION.
 CC -!- THE C-TERMINAL OF THE HEAVY CHAIN BINDS TO GANGLIOSIDE RECEPTORS.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27 (ZINC
 CC METALLOPROTEASE); ALSO KNOWN AS THE TETANUS/BOTULINUM NEUROTOXIN
 CC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X04436; G40770; -
 DR EMBL; M12739; G144921; -
 DR EMBL; X05214; G40774; -
 DR PIR; A25689; BTCLIN.
 DR PDB; 1AF9; 29-APR-98.
 DR PROSITE; PS00142; ZINC-PROTEASE; 1.
 KW NEUROTOXIN; TRANSMEMBRANE; HYDROLASE; METALLOPROTEASE; ZINC; PLASMID;
 FT 3D-STRUCTURE.
 FT INIT_MET 0 0
 FT CHAIN 1 456
 FT CHAIN 457 1314
 FT CHAIN 1315 1314
 FT METAL 232 232
 FT ACT_SITE 233 233
 FT METAL 236 236
 FT METAL 226 246
 FT TRANSMEM 569 589
 FT DISULFID 438 466
 FT DISULFID 1076 1092
 SQ SEQUENCE 1314 AA; 150550 MW; 26190E3E CRC32;
 Query Match 100.0%; Score 104; DB 1; Length 1314;
 Best Local Similarity 100.0%; Pred. No. 3.53e-12;
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 829 QYKANSKFIGITEL 843
 QY 1 QYKANSKFIGITEL 15
 RESULT 2
 ID RAD_MYCCA STANDARD; PRT; 213 AA.
 AC P10251;
 DT 01-MAR-1989 (REL. 10, CREATED)
 DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE ADENYLATE KINASE (EC 2.7.4.3) (ATP-AMP TRANSPHOSPHORYLASE).
 GN ADK.
 OS MYCOPLASMA CAPRICOLUM.
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; MOLLICUTES;
 OC CAPRICOLUM GROUP.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 27343 / KID;
 RX MEDLINE; 88142549.
 RA OKUBO S., MUTO A., KAWAUCHI Y., YAMAO F., OSANA S.;
 RT "The ribosomal protein gene cluster of Mycoplasma capricolum.";
 RL MOL. GEN. GENET. 210:314-322(1987).
 CC -!- FUNCTION: THIS SMALL UBIQUITOUS ENZYME IS ESSENTIAL FOR
 CC MAINTENANCE AND CELL GROWTH.
 CC -!- CATALYTIC ACTIVITY: ATP + AMP = ADP + ADP.
 CC -!- SUBUNIT: MONOMER.
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -!- SIMILARITY: BELONGS TO THE ADENYLATE KINASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X06414; G44229; -
 DR PIR; S02851; KIYMC.
 DR PROSITE; PS00113; ADENYLATE_KINASE; 1.
 DR PFAM; PF00406; adenylatekinase; 1.
 DR HSSP; P27142; 1ZIO.
 KW TRANSFERASE; KINASE; ATP-BINDING.
 FT NP_BIND 7 15
 FT SEQUENCE 213 AA; 24616 MW; 8D5C4B03 CRC32;
 Query Match 60.6%; Score 63; DB 1; Length 213;
 Best Local Similarity 66.7%; Pred. No. 8.08e-02;
 Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 Db 182 DYFKNSKFEI 193
 QY 1 QYKANSKFEI 12
 RESULT 3
 ID V120.HSVSA STANDARD; PRT; 899 AA.
 AC Q01055;
 DT 01-APR-1993 (REL. 25, CREATED)
 DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
 DE CAPSID ASSEMBLY PROTEIN 63.
 GN 63 OR EERF1.
 OS HERPESVIRUS SAIMIRI (STRAIN 11).
 OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
 OC GAMMAHERPESVIRINAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE; 92333688.
 RA ALBRECHT J.-C., NICHOLAS J., BILLER D., CAMERON K.R., BIESINGER B.,
 RA NEWMAN C., WITTMANN S., CRAXTON M.A., COLEMAN H., FLECKENSTEIN B.,
 RA HONESS R.W.;

```
RT "Primary structure of the herpesvirus saimiri genome.";
RN J. VIROL. 66:5047-5058(1992).
RP SEQUENCE FROM N.A.
RX MEDLINE; 92230228.
RA NICHOLAS J., CAMERON K.R., COLEMAN H., NEWMAN C., HONNESS R.W.;
RT "Analysis of nucleotide sequence of the rightmost 43 kbp of
RI herpesvirus saimiri (HVS) L-DNA: general conservation of genetic
RT organization between HVS and Epstein-Barr virus.";
RL VIROLOGY 188:296-310(1992).
CC -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL37,
CC EBV-1 23, EBV BOF1, VZV 21, HVS-1 63, AND HCMV UL47.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X63436; G60384; -.
DR EMBL; M86409; G330992; -.
DR PIR; G36812; G36812.
KW CAPSID ASSEMBLY.
SQ SEQUENCE 899 AA; 103350 MW; C2D70154 CRC32;

Query Match 50.6%; Score 63; DB 1; Length 899;
Best Local Similarity 50.0%; Pred. No. 8,08e-02;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 124 QYTSNATITGLSE 137
QY 1 QYKANSKFIGITE 14
|||||:|:|:|:|:|

RESULT 4
ID RIBF_SVNY3 STANDARD; PRI: 284 AA.
AC P73651;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE RIBOFLAVIN KINASE (EC 2.7.1.26) (FLAVOKINASE) / FMN
DE ADENYLITYLTRANSFERASE (EC 2.7.7.2) (FAD PYROPHOSPHORYLASE) (FAD
DE SYNTHETASE)
GN RIBF OR SLR1882.
OS SYNECHOCYSTIS SP. (STRAIN PCC 6803).
OC BACTERIA; CYANOBACTERIA; CHROCOCCALES; SYNECHOCYSTIS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97061201.
RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
RA MIYAJIMA T., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K.,
RA OKUMURA S., SHIMPO S., TAKEUCHI C., WADA T., WATANABE A.,
RA YAMADA M., YASUDA M., TABATA S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA RES. 3:109-136(1996).
CC -!- CATALYTIC ACTIVITY: ATP + FMN = DIPHOSPHATE + FAD.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D90908; G165277; -.
DR TRANSFERASE; NUCLEOTIDYLTRANSFERASE; MULTIFUNCTIONAL ENZYME.
KW
```

```
SQ SEQUENCE 284 AA; 31380 MW; A4BFA0C CRC32;

Query Match 59.6%; Score 62; DB 1; Length 284;
Best Local Similarity 33.3%; Pred. No. 1.34e-01;
Matches 5; Conservative 8; Mismatches 2; Indels 0; Gaps 0;

Db 248 HYLRPETKFAQLDQ 262
QY 1 QYKANSKFIGITEL 15
|||||:|:|:|:|:|

RESULT 5
ID ACPD_HAEIN STANDARD; PRT: 194 AA.
AC P43013;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE ACYL CARRIER PROTEIN PHOSPHODIESTERASE (ACP PHOSPHODIESTERASE).
GN ACPD OR H11366.
OS HAEMOPHILUS INFLUENZAE.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURELLACEAE;
OC HAEMOPHILUS.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-RD;
RA CHANDLER M.S., SMITH R.A.;
RL SUBMITTED (FEB-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN-RD / KW20;
RX MEDLINE; 95350630.
RA FLEISCHMANN R.D., WHITE O., ADAMS M.D., CLAYTON R.A., KIRKNESS E.F.,
RA KERLAVAGE A.R., BULT C.J., TOMB J.-F., DOUGHERTY B.A., MERRICK J.M.,
RA MCKENNEY K., SUTTON G., FITZHUGH W., FIELDS C.A., GOCAYNE J.D.,
RA SCOTT J.D., SHIRLEY R., LIU L.-I., GLODEK A., KELLEY J.M.
RA WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDELOM E., COTTON M.D.,
RA UTTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDEK D.M., BRANDON R.C.,
RA FINE L.D., FRITCHMAN J.L., FUHRMANN J.L., GEOGHAGEN N.S.M.,
RA GNEHM C.L., McDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,
RA VENTER J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RL SCIENCE 269:496-512(1995).
CC -!- FUNCTION: CONVERTS HOLO-ACP TO APO-ACP BY HYDROLYTIC CLEAVAGE OF
CC THE PHOSPHOPANTHEINE RESIDUE FROM ACP. ITS PHYSIOLOGICAL
CC FUNCTION IS NOT CLEAR (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U20964; G687791; -.
DR EMBL; U32816; G1574198; -.
DR TIGR; H11366; -.
KW HYDROLASE.
SQ SEQUENCE 194 AA; 21208 MW; B4D866EF CRC32;

Query Match 58.7%; Score 61; DB 1; Length 194;
Best Local Similarity 53.3%; Pred. No. 2.20e-01;
Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 147 QYKXSLGFTGITHV 161
QY 1 QYKANSKFIGITEL 15
|||||:|:|:|:|:|

RESULT 6
ID GGT_PIG STANDARD; PRT: 568 AA.
AC P20735;
```

01-FEB-1991 (REL. 17, CREATED)
01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
GAMMA-GLUTAMYLTRANSPEPTIDASE PRECURSOR (EC 2.3.2.2) (GAMMA-GLUTAMYLTRANSFERASE) (GGT).
GGT.
SUS SCROFA (PIG).
EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
ARTIODACTYLA; SUIFORMES; SUINA; SUIDAE; SUS.
[1]
SEQUENCE FROM N.A.
TISSUE-BRAIN;
MEDLINE; 89377838.
PAPANIKOPOULOU A., FREY A., GASSEN H.G.;
"Cloning and expression of gamma-glutamyl transpeptidase from isolated porcine brain capillaries";
EUR. J. BIOCHEM. 183:693-698(1989).
CC -1- CATALYTIC ACTIVITY: (S-L-GLUTAMYL)-PEPTIDE + AN AMINO ACID = PEPTIDE + S-L-GLUTAMYL-AMINO ACID.
CC -1- PATHWAY: GGT PLAYS A KEY ROLE IN THE GAMMA-GLUTAMYL CYCLE, A PATHWAY FOR THE SYNTHESIS AND DEGRADATION OF GLUTATHIONE.
CC -1- SUBUNIT: THIS ENZYME CONSISTS OF TWO POLYPEPTIDE CHAINS, WHICH ARE SYNTHESIZED IN PRECURSOR FORM FROM A SINGLE POLYPEPTIDE.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
CC -1- SIMILARITY: STRONG, TO OTHER GGT ENZYMES AND TO GL-7ACA ACYLASES.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; Z46922; G500819; -;
EMBL; X16533; G502909; -;
PIR; S05532; S05532.
DR DR PROSITE; PS00462; G.GLU_TRANSPEPTIDASE; 1.
DR DR PFAM; PF01019; G.glu_transpept; 1.
KW TRANSFERASE; ACYLTRANSFERASE; ZMOGEN; GLYCOPROTEIN; TRANSMEMBRANE;
KW GLUTATHIONE BIOSYNTHESIS; SIGNAL-ANCHOR.
FT CHAIN 1 379 HEAVY CHAIN.
FT CHAIN 380 568 LIGHT CHAIN.
FT FT TRANSMEM 5 26 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (PROBABLE).
FT FT DOMAIN 27 568 LUMENAL (POTENTIAL).
FT FT BINDING 453 453 GAMMA-GLUTAMYL (POTENTIAL).
FT FT CARBOHYD 94 94 POTENTIAL.
FT FT CARBOHYD 119 119 POTENTIAL.
FT FT CARBOHYD 229 229 POTENTIAL.
FT FT CARBOHYD 296 296 POTENTIAL.
FT FT CARBOHYD 336 336 POTENTIAL.
FT FT CARBOHYD 343 343 POTENTIAL.
FT FT CARBOHYD 427 427 POTENTIAL.
FT FT CARBOHYD 510 510 POTENTIAL.
SQ SEQUENCE 568 AA; 61315 MW; D5EEE3C4 CRC32;

Query Match 57.7%; Score 60; DB 1; Length 568;
Best Local Similarity 42.9%; Pred. No. 3.61e-01;
Matches 6; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Db 531 HYIQDASTFGIVGVQ 544
:|:| | | | :
QY 1 QYIKANSKFGITE 14

RESULI 7
ID VG84_BPML5 STANDARD; PRT; 66 AA.
AC Q05301;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
DE GENE 84 PROTEIN (GP84).

```

Best Local Similarity 80.0%; Pred.No. 5.87e-01;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 266 DSKFIGIGEL 275
QY 6 NSKFIGITEL 15

-----
RESULT 9
ID RIBF MYCPN STANDARD; PRT; 269 AA.
AC P75537;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DE PUTATIVE RIBOFLAVIN KINASE (EC 2.7.1.26) (FLAVOKINASE) / FMN
DE ADENYLTRANSFERASE (EC 2.7.7.2) (FAD PYROPHOSPHORYLASE) (FAD
DE SYNTHETASE).
GN RIBF.
OS MYCOPLASMA PNEUMONIAE.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; MOLLICUTES;
OC MYCOPLASMATACEAE; MYCOPLASMA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE; 97105885.
RA HIMMELREICH R., HILBERT H., PLAGENS H., PIRKL E., LI B.-C.,
RA HERRMANN R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae";
RL NUCLEIC ACIDS RES. 24:4420-4449(1996).
CC -|- CATALYTIC ACTIVITY: ATP + RIBOFLAVIN = ADP + FMN.
CC -|- CATALYTIC ACTIVITY: ATP + FMN = DIPHOSPHATE + FAD.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
Db EMBL; AE000062; G1674379; -
KW TRANSFERASE; NUCLEOTIDYLTRANSFERASE; MULTIFUNCTIONAL ENZYME.
SQ SEQUENCE 269 AA; 30435 MW; 07AF8D7C CRC32;

Query Match 54.8%; Score 57; DB 1; Length 269;
Best Local Similarity 40.0%; Pred.No. 1.53e-00;
Matches 6; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Db 235 QFIRPQKFGVQAL 249
QY 1 QYIKANSKFIGITEL 15

-----
RESULT 10
ID DPOB_XENLA STANDARD; PRT; 333 AA.
AC Q57383;
DT 15-DEC-1998 (REL. 37, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DE DNA POLYMERASE BETA (EC 2.7.7.7).
GN POLB.
OS XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC EUKARYOTA; METAEOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC MESOBATRACHIA; PIPOIDEA; PIPIIDAE; XENOPODINAE; XENOPUS.
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE; 98151235.
RA REICHENBERGER S., PFEIFFER P.;
RT "Cloning, purification and characterization of DNA polymerase beta
RT from Xenopus laevis -- studies on its potential role in DNA-end
RT joining.";
RL EUR. J. BIOCHEM. 251:81-90(1998).

-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC EMBL; Y15732; E1198727; -
DR PROSITE; PS00522; DNA-POLYMERASE X; 1.
KW TRANSFERASE; DNA-DIRECTED DNA POLYMERASE; DNA REPLICATION.
FT INIT_MET 0 0 BY SIMILARITY.
FT ACT_SITE 182 182 INVOLVED IN PRIMER BINDING
FT (BY SIMILARITY).
FT ACT_SITE 189 189 INVOLVED IN PRIMER BINDING
FT (BY SIMILARITY).
FT ACT_SITE 191 191 INVOLVED IN PRIMER BINDING
FT (BY SIMILARITY).
FT ACT_SITE 191 191 INVOLVED IN PRIMER BINDING
FT (BY SIMILARITY).
SQ SEQUENCE 333 AA; 38162 MW; 6948CECC CRC32;

Query Match 54.8%; Score 57; DB 1; Length 333;
Best Local Similarity 38.5%; Pred.No. 1.53e-00;
Matches 5; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

Db 228 VKGDTKFMGVQQL 240
QY 3 IKANSKFIGITEL 15

-----
RESULT 11
ID MSPL_PLAFK STANDARD; PRT; 1630 AA.
AC P04932;
DT 13-AUG-1987 (REL. 05, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE (PMMSA) (P190).
GN MSP-1.
OS PLASMODIUM FALCIPARUM (ISOLATE K1 / THAILAND).
OC EUKARYOTA; ALVEOLATA; APICOMPLEXA; HAEMOSPORIDA; PLASMODIUM.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 86136024.
RA MACKAY M., GOMAN M., BONE N., HYDE J.E., SCAIFE J., CERTA U.,
RA STUNNENBERG H., BUJARD H.;
RT "Polymorphism of the precursor for the major surface antigens of
RT Plasmodium falciparum merozoites: studies at the genetic level.";
RL EMBO J. 4:3823-3829(1985).
RN [2]
RP REVISIONS, SEQUENCE FROM N.A.
RA PAN W., TOLLE R., BUJARD H.;
RL SUBMITTED (JUN-1995) TO EMBL/GENEBANK/DBJ DATA BANKS.
CC -|- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
(POTENTIAL).
CC -|- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KD, 42
KD AND 19 KD ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----

```

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; X03371; G929798; -
 CC DR PIR; A25120; SAZOKI.1
 CC DR PFAM; PF000008; EGF; 1.
 CC KW MALARIA; MEROZOITE; POLYPROTEIN; REPEAT; SIGNAL; GLYCOPROTEIN;
 KW TRANSMEMBRANE; GPI-ANCHOR.

FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 1630 MEROZOITE SURFACE PROTEIN 1.
 FT DOMAIN 67 84 TRIPEPTIDE SG(TP) REPEAT.
 FT TRANSMEM 1614 1630 MEMBRANE ANCHOR.
 FT CARBOHYD 97 97 POTENTIAL.
 FT CARBOHYD 259 259 POTENTIAL.
 FT CARBOHYD 755 755 POTENTIAL.
 FT CARBOHYD 759 759 POTENTIAL.
 FT CARBOHYD 774 774 POTENTIAL.
 FT CARBOHYD 835 835 POTENTIAL.
 FT CARBOHYD 911 911 POTENTIAL.
 FT CARBOHYD 955 955 POTENTIAL.
 FT CARBOHYD 1049 1049 POTENTIAL.
 FT CARBOHYD 1156 1156 POTENTIAL.
 FT CARBOHYD 1165 1165 POTENTIAL.
 FT CARBOHYD 1436 1436 POTENTIAL.
 FT CARBOHYD 1517 1517 POTENTIAL.
 SQ SEQUENCE 1630 AA; 187289 MW; DD2F8628 CRC32;

Query Match 54.8%; Score 57; DB 1; Length 1630;
 Best Local Similarity 46.2%; Pred. No. 1.53e+00;
 Matches 6; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

Db 1464 FKNNNFVGIADL 1476
 QY 3 IKANSXFIGITEL 15

RESULT 12
 ID MSP1_PLAFW STANDARD; PRI; 1639 AA.
 AC P04933;
 DT 13-AUG-1987 (REL. 05, CREATED)
 DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
 DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
 DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
 DE (PMSEA) (P195).
 GN MSP-1.
 OS PLASMODIUM FALCIPARUM (ISOLATE WELLCOME).
 OC EUKARYOTA; ALVEOLATA; APICOMPLEXA; HAEMOSPORIDA; PLASMODIUM.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 86014355.
 RA HOLDER A.A.; LOCKYER M.J.; ODINK K.G.; SANDHU J.S.; RIVEROS-MORENO V.,
 RA NICHOLLS S.C.; HILLMAN Y.; DAVEY L.S.; TIZARD M.L.V.; SCHWARZ R.T.,
 RA FREEMAN R.R.;
 RT "Primary structure of the precursor to the three major surface
 RT antigens of Plasmodium falciparum merozoites.";
 RL NATURE 317:270-273(1985).
 RN [2]
 RP REVISIONS.
 RA HOLDER A.A.;

CC SUBMITTED (MAR-1991) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
 CC (POTENTIAL).
 CC -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KD, 42
 CC KD AND 19 KD ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
 CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC or send an email to license@isb-sib.ch).

CC EMBL; X02919; G9865; -
 CC DR PIR; A24594; A24594.
 CC DR PFAM; PF000008; EGF; 1.
 CC KW MALARIA; MEROZOITE; POLYPROTEIN; REPEAT; SIGNAL; GLYCOPROTEIN;
 KW TRANSMEMBRANE; GPI-ANCHOR.

FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 1639 MEROZOITE SURFACE PROTEIN 1.
 FT CARBOHYD 116 116 POTENTIAL.
 FT CARBOHYD 268 268 POTENTIAL.
 FT CARBOHYD 764 764 POTENTIAL.
 FT CARBOHYD 768 768 POTENTIAL.
 FT CARBOHYD 783 783 POTENTIAL.
 FT CARBOHYD 844 844 POTENTIAL.
 FT CARBOHYD 920 920 POTENTIAL.
 FT CARBOHYD 964 964 POTENTIAL.
 FT CARBOHYD 1058 1058 POTENTIAL.
 FT CARBOHYD 1165 1165 POTENTIAL.
 FT CARBOHYD 1174 1174 POTENTIAL.
 FT CARBOHYD 1445 1445 POTENTIAL.
 FT CARBOHYD 1526 1526 POTENTIAL.
 SQ SEQUENCE 1639 AA; 187618 MW; F0860D6A CRC32;

Query Match 54.8%; Score 57; DB 1; Length 1639;
 Best Local Similarity 46.2%; Pred. No. 1.53e+00;
 Matches 6; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

Db 1473 FKNNNFVGIADL 1485
 QY 3 IKANSXFIGITEL 15

RESULT 13
 ID RIBC_BACSU STANDARD; PRI; 316 AA.
 AC P54575; P70987;
 DT 01-OCT-1996 (REL. 34, CREATED)
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE RIBOFLAVIN KINASE (EC 2.7.1.26) (FLAVOKINASE) / FMN
 DE ADENYLYLTRANSFERASE (EC 2.7.7.2) (FAD PYROPHOSPHORYLASE) (FAD
 DE SYNTHETASE).
 GN RIBC.
 OS BACILLUS SUBTILIS.
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
 OC BACILLUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RA GUSAROV I.V.; YOMANTAS Y.I.; KOZLOV Y.I.; KRENEVA R.A.,
 RA PEROMOV D.A.;
 RL SUBMITTED (JAN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RA COQUARD D.; HUECAS M.; OTT M.; VAN DIJL J.; VAN LOON A.; HOHMANN H.;
 RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC -!- CATALYTIC ACTIVITY: ATP + RIBOFLAVIN = ADP + FMN.
 CC -!- CATALYTIC ACTIVITY: ATP + FMN = DIPHOSPHATE + FAD.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; X95312; E219687; -
 CC DR EMBL; E20835; E269877; -
 CC DR EMBL; Z99112; E1185258; -
 CC SUBMITT; BG11495; RIBC.
 KW TRANSFERASE; NUCLEOTIDYLTRANSFERASE; MULTIFUNCTIONAL ENZYME.

```

OC RHABDITINA; RHABDITOIDEA; RHABDITIIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA WHITE S.;
RL SUBMITTED (NOV-1995) TO EMBL/GENEBANK/DDBJ DATA BANKS.
CC -1- CATALYTIC ACTIVITY: CITRATE = CIS-ACONITATE + H(2)O.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE ACONITASE/TPM ISOMERASE FAMILY.
-----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: Z66567; E1351089; -.
CC WORMPEP; ZK455.1; CE03812.
CC PROSITE; PS00450; ACONITASE_1; 1.
CC PROSITE; PS01244; ACONITASE_2; 1.
CC PFAM; PF00330; aconitase; 1.
CC PFAM; PF00694; Aconitase.C; 1.
CC LYASE; TRICARBOXYLIC ACID CYCLE; IRON-SULFUR; 4FE-4S. (BY
CC METAL 436 436 IRON-SULFUR (4FE-4S) (BY SIMILARITY)).
CC FT METAL 502 502 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
CC FT METAL 505 505 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
CC SQ SEQUENCE 887 AA; 96560 MW; 226EF357 CRC32;
-----
Query Match 53.8%; Score 56; DB 1; Length 887;
Best Local Similarity 35.7%; Pred.No. 2.44e+00;
Matches 5; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

Db 334 QYLKSVGMFVNFTD 347
||:|: |:|: |:|:
Qy 1 QYIKANSKFIGITE 14

Search completed: Tue Aug 17 15:48:46 1999
Job time : 20 secs.

```

This Page Blank (uspto)

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Tue Aug 17 15:49:04 1999; Maspar time 6.15 Seconds
Tabular output not generated.
133.075 Million cell updates/sec

Title: >US-09-049-847-1
Description: (1-15) from US09049847.pep
Perfect Score: 104
Sequence: 1 QYIKANSKFIGITEL 15

Scoring table: PAM 150
Gap 15

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: spiremb19

1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human
5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organelle
9:sp-phage 10:sp-plant 11:sp-rodent 12:sp_unclassified
13:sp-vertebrate 14:sp_virus

Statistics: Mean 26.069; Variance 32.264; scale 0.808

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	62	59.6	598	2	YEEB PROTEIN.	2.21e-01
2	61	58.7	349	3	FISSION YEAST (FRAGMEN	3.62e-01
3	59	56.7	1048	5	HYPOTHETICAL PROTEIN (9.52e-01
4	59	56.7	3119	5	PFG377.	9.52e-01
5	58	55.8	131	7	MHC CLASS II BETA CHAI	1.53e+00
6	58	55.8	357	2	YOKG PROTEIN.	1.53e+00
7	58	55.8	357	9	YOKG PROTEIN.	1.53e+00
8	58	55.8	423	2	TYPE IIS RESTRICTION E	1.53e+00
9	57	54.8	400	5	MEROZOITE SURFACE ANTI	2.46e+00
10	57	54.8	421	5	EG-34F3.5 PROTEIN.	2.46e+00
11	57	54.8	424	2	HYPOTHETICAL 45.1 KD P	2.46e+00
12	57	54.8	539	5	MAJOR MEROZOITE SURFAC	2.46e+00
13	57	54.8	539	5	MAJOR MEROZOITE SURFAC	2.46e+00
14	57	54.8	539	5	MAJOR MEROZOITE SURFAC	2.46e+00
15	57	54.8	539	5	MAJOR MEROZOITE SURFAC	2.46e+00
16	57	54.8	539	5	MAJOR MEROZOITE SURFAC	2.46e+00
17	57	54.8	539	5	MAJOR MEROZOITE SURFAC	2.46e+00
18	57	54.8	539	5	MAJOR MEROZOITE SURFAC	2.46e+00
19	57	54.8	1333	5	BLASTOPIA POLYPROTEIN.	2.46e+00
20	56	53.8	227	5	COMPLETE GENE MAP OF P	3.91e+00

21	56	53.8	271	9	048471	COMPLETE NUCLEOTIDE SE	3.91e+00
22	56	53.8	381	1	059512	38AA LONG HYPOTHETICA	3.91e+00
23	56	53.8	601	2	P94880	OLIGOPEPTIDASE.	3.91e+00
24	56	53.8	1018	5	017874	F46F6.2 PROTEIN.	3.91e+00
25	55	52.9	131	7	046866	MHC CLASS II BETA CHAI	6.19e+00
26	55	52.9	249	7	P79565	MHC CLASS II BETA CHAI	6.19e+00
27	55	52.9	290	2	050869	CHEMOTACTIC RESPONSE R	6.19e+00
28	55	52.9	382	2	P94415	RESPONSE-REGULATOR ASP	6.19e+00
29	55	52.9	436	5	P91071	COSMID C18B10.	6.19e+00
30	55	52.9	447	11	P70680	ALPHA-1,3-MANNOSYL-GLY	6.19e+00
31	55	52.9	458	5	044506	F42G8.9 PROTEIN.	6.19e+00
32	55	52.9	509	5	027482	SIMILAR TO CYTOCHROME	6.19e+00
33	54	51.9	108	2	054172	PUTATIVE CYCLASE.	6.19e+00
34	54	51.9	345	2	030883	ERPX PROTEIN.	9.74e+00
35	54	51.9	413	3	P78759	FISSION YEAST (FRAGMEN	9.74e+00
36	54	51.9	446	10	P93632	GLOSSY15.	9.74e+00
37	54	51.9	509	5	027499	SIMILAR TO CYTOCHROME	9.74e+00
38	54	51.9	1102	10	049552	UV-DAMAGED DNA-BINDING	9.74e+00
39	53	51.0	130	1	059432	130AA LONG HYPOTHETICA	1.52e+01
40	53	51.0	188	2	044863	ORFB (ORF2).	1.52e+01
41	53	51.0	188	2	025607	HYPOTHETICAL 21.2 KD P	1.52e+01
42	53	51.0	188	2	045022	REPEATED DNA ELEMENT,	1.52e+01
43	53	51.0	188	2	044870	PLASMID, ORFA. B. C. D	1.52e+01
44	53	51.0	294	2	025678	ATP-BINDING PROTEIN (Y	1.52e+01
45	53	51.0	652	2	068071	DNA GYRASE SUBUNIT B (1.52e+01

ALIGNMENTS

RESULT 1
ID 034469 PRELIMINARY; PRT; 598 AA.
AC 034469;
DT 01-JAN-1998 (TREMELREL. 05, CREATED)
DT 01-JAN-1998 (TREMELREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMELREL. 08, LAST ANNOTATION UPDATE)
DE YEEB PROTEIN.
GN YEEB.
OS BACILLUS SUBTILIS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE: 98044033.
RA KUNST F., OGAWARA N., MOSZER I., ALBERTINI A.M., AILONI G.,
RA AZEVEDO V., BERTIERO M.G., BESSIERES P., BOLOTIN A., BORCHERT S.,
RA BORRIS R., BOURSIER L., BRANS A., BRAUN M., BRIGHELL S.C., BRON S.,
RA BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M.,
RA CHOI S.K., CODANI J.J., CONNERTON I.F., CUMMINGS N.J., DANIEL R.A.,
RA DENIZOT F., DEVINE K.M., DUSTERHOFT A., EHRLICH S.D., EMERSON P.T.,
RA ENRIAN K.D., ERRINGTON J., FABRET C., FERRARI E., FOULGER D.,
RA FRITZ C., FUJITA M., FUJITA Y., FUMA S., GALIZZI A., GALLERON N.,
RA GHIM S.Y., GLASER P., GOFFEAU A., GOLIGHTLY E.J., GRANDI G.,
RA GUISEPPI G., GUY B.J., HAGA K., HAIECH J., HARWOOD C.R., HENAUT A.,
RA HILBERT H., HOLTSAPPEL S., HOSONO S., HULLO M.F., IATAI M., JONES L.,
RA JORTS B., KARAMATA D., KASHARA Y., KLAERR-BLANCHARD M., KLEIN C.,
RA KOBAYASHI Y., KOTTER P., KONINGSTEIN G., KROGH S., KUMANO M.,
RA KURITA K., LAPIDUS A., LARDINOIS S., LAUBER J., LAZAREVIC V.,
RA LEE S.M., LEVINE A., LIU H., MASUDA S., MAUEL C., MEDIGUE C.,
RA MEONE D., MELLADO R.P., MIZUNO M., MOESTL D., NAKAI S., NOBACK M.,
RA NOONE N., O'REILLY M., OGAWA K., OGILWRA A., ODEGA B., PARK S.H.,
RA PARRO V., POHL T.M., PORTELELLA D., PORWOLLIK S., PRESCOTT A.M.,
RA PRESECAN E., PUJIC P., PURNELLE B., RAPOPORT G., REY M., REYNOLDS S.,
RA RIEGER M., RIVOLTA C., ROCHA E., ROCHE B., ROSE M., SADALE Y.,
RA SATO T., SCANLAN E., SCHLEICH S., SCHROTER R., SCOFFONE F.,
RA SEKIGUCHI J., SEKOWSKA A., SEROR S.J., SEROR P., SHIN B.S., SOLDI B.,
RA SOROKIN A., TACCONI E., TAKAGI T., TAKAHASHI H., TAKEMARU K.,
RA TOSUCHI M., TANAKA T., TERPSTRA P., TOGNONI A.,
RA TOSATO V., UCHIYAMA S., VANDENBOL M., VANNIER F., VASSAROTTI A.,
RA VIARI A., WAMEBUTT R., WEDLER E., WEDLER H., WEITZNEGER I.,
RA WINTERS P., WIPAT A., YAMAMOTO H., YAMANE K., YASUMOTO K., YATA K.,
RA YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E., YOSHIKAWA H., DANCHIN A.,
RT "The complete genome sequence of the gram-positive bacterium Bacillus


```

RL  SUBMITTED (AUG-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR  EMBL: AF020713; G3025485; -.
SQ  SEQUENCE 357 AA; 40742 MW; 56ABC286 CRC32;

Query Match          55.8%; Score 58; DB 9; Length 357;
Best Local Similarity 35.78; Pred. No. 1.53e+00;
Matches 5; Conservative 7; Mismatches 2; Indels

Db 165 FLOGGNFVIGVTQL 178
QY 2 YIKANSKFIGITEL 15

RESULT 8
ID Q25919 PRELIMINARY; PRT; 423 AA.
AC Q25919;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DI 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE TYPE IIS RESTRICTION ENZYME R PROTEIN (MBCIIR).
GN HPI366.
OS HELICOBACTER PYLORI (CAMPYLOBACTER PYLORI).
OC BACTERIA; PROTEOBACTERIA; EPSILON SUBDIVISION; HELICOBACTERIA;
OC HELICOBACTER.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-26695;
RX MEDLINE; 97394467.
RA TOMB J.-F., WHIE O., KERLAVAGE A.R., CLAYTON R.A., SUTTORP
RA FLEISCHMANN R.D., KETCHUM K.A., KLENK H.-P., GILL S., DOUG
RA NELSON B., RICHARDENBUSH J., ZHOU L., KIRKNESS E.F., PETERS
RA LOFTUS K., RICHARDSON D., DODSON R., KEALAK H.G., GLODEK
RA MCKENNEY K., FITZGERALD L.M., LEE N., ADAMS M.D., HICKEY
RA BERG D.E., GOCAYNE J.D., UTTERBACK T.R., PETERSON J.D., K
RA COTTON M.D., WEIDMAN J.M., FUJII C., BOWMAN C., WATTHEY L
RA HAYES W.S., BORODOVSKY M., KARP P.D., SMITH H.C., FRASER
RA VENTER J.C.;
RT "The complete genome sequence of the gastric pathogen Hel
RT pylori [published erratum appears in Nature 1997 Sep
RT 25; 389(5649):412].";
RT NATURE 388:539-547(1997).
DR EMBL: AE000637; G2314537; -.
DR TIGR: HPI366; -.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 423 AA; 50047 MW; BD8FE438 CRC32;

Query Match          55.8%; Score 58; DB 2; Length 423;
Best Local Similarity 50.08; Pred. No. 1.53e+00;
Matches 7; Conservative 4; Mismatches 3; Indels

Db 140 YIQSNIKYISLTPL 153
QY 2 YIKANSKFIGITEL 15

RESULT 9
ID Q03999 PRELIMINARY; PRT; 400 AA.
AC Q03999;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DI 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE MEROZOITE SURFACE ANTIGEN PRECURSOR 1 (PMMSA) (FRAGMENT).
GN MSP1.
OS PLASMODIUM FALCIPARUM.
OC EUKARYOTA; ALVEOLATA; APICOMPLEXA; HAEMOSPORIDA; PLASMODIUM
OC [1]
RP SEQUENCE FROM N.A.
RA BLACKMAN M.J., LING I.T., NICHOLLS S.C., HOLDER A.A.;
RL SUBMITTED (XXX-1991) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83K.
CC ANIGENS WHICH ARE THE MAJOR SURFACE ANIGENS OF MEROC
CC THE MATURATION TAKE PLACE DURING SCHIZONT.
CC -!- SUBCELLULAR LOCATION: COULD BE ATTACHED TO THE MEMBRAN

```

CC GPI-ANCHOR.
 DR EMBL; M64681; GEG; 1.
 DR PFAM; PF00008; EGF; 1.
 KW MALARIA: MEROZOITE; POLYPROTEIN; REPEAT; SIGNAL; GLYCOPROTEIN;
 KW TRANSMEMBRANE; GPI-ANCHOR.

FT NON_TER 1
 FT SIGNAL <1 24
 FT CHAIN 25 400
 FT CHAIN 25 285
 FT CHAIN 287 400
 FT TRANSMEM 333 400
 SQ SEQUENCE 400 AA; 45824 MW; 0E131D8C CRC32;

Query Match 54.8%; Score 57; DB 5; Length 400;
 Best Local Similarity 46.2%; Pred. No. 2.46e+00;
 Matches 6; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

Db 234 FKNNNFVGIADL 246
 QY 3 IKANSKFIGITEL 15

RESULT 10
 ID O77436 PRELIMINARY; PRT; 421 AA.
 AC O77436;
 DT 01-NOV-1998 (TREMELREL. 08, CREATED)
 DT 01-NOV-1998 (TREMELREL. 08, LAST SEQUENCE UPDATE)
 DE 01-NOV-1998 (TREMELREL. 08, LAST ANNOTATION UPDATE)
 DE EG:34F3.5 PROTEIN.
 GN EG:34F3.5.
 OS DROSOPHILA MELANOGASTER (FRUIT FLY).
 OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
 OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
 OC DROSOPHILIDAE; DROSOPHILA.
 RN [1]

SEQUENCE FROM N.A.
 RA CATHERINE SALLES, PHILIPPE VALENTI, ARETI DARLAMITSOV,
 RA NADINE HENDERSON, LORNA CAMPBELL, DAVID GLOVER;
 RT "Sequencing the distal x chromosome of Drosophila melanogaster."
 RL SUBMITTED (SEP-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]

SEQUENCE FROM N.A.
 RA BENOS P.;
 RL SUBMITTED (OCT-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; AL031583; E1321021; -.
 SQ SEQUENCE 421 AA; 48676 MW; 255FB6CF CRC32;

Query Match 54.8%; Score 57; DB 5; Length 421;
 Best Local Similarity 42.9%; Pred. No. 2.46e+00;
 Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Db 198 YIRNERFNFMTL 211
 QY 2 YKANSKFIGITEL 15

RESULT 11
 ID O86555 PRELIMINARY; PRT; 424 AA.
 AC O86555;
 DT 01-NOV-1998 (TREMELREL. 08, CREATED)
 DT 01-NOV-1998 (TREMELREL. 08, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMELREL. 08, LAST ANNOTATION UPDATE)
 DE HYPOTHETICAL 45.1 KD PROTEIN.
 GN SC1F2.18.

OS STREPTOMYCES COELICOLOR.
 OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
 OC ACTINOMYCETALES; STREPTOMYCINAE; STREPTOMYCETACEAE; STREPTOMYCES.
 RN [1]

SEQUENCE FROM N.A.
 RA STRAIN 76(2);
 RA SEEGER K.J.; HARRIS D.;
 RL SUBMITTED (AUG-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA PARKHILL J.; BARRELL B.G.; RAJANDREAM M.A.;
 RL SUBMITTED (AUG-1998) TO EMBL/GENBANK/DBJ DATA BANKS.

RN [3]
 RP SEQUENCE FROM N.A.

RC STRAIN-A3(2);
 RC MEDLINE; 97000351.
 RA REDENBACH M.; KIESER H.M.; DENAPATE D.; EICHNER A.; CULLUM J.;
 RA KINASHI H.; HOPWOOD D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb streptomyces coelicolor A3(2) chromosome."
 RL MOL. MICROBIOL. 21:77-96(1996).
 DR EMBL; AL031350; E1316910; -.
 KW HYPOTHETICAL PROTEIN.

SQ SEQUENCE 424 AA; 45064 MW; B4BC2A68 CRC32;

Query Match 54.8%; Score 57; DB 2; Length 424;
 Best Local Similarity 46.7%; Pred. No. 2.46e+00;
 Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Db 358 QYKAGDTLYGATDL 372
 QY 1 QYKANSKFIGITEL 15

RESULT 12
 ID Q25981 PRELIMINARY; PRT; 539 AA.
 AC Q25981;
 DT 01-NOV-1996 (TREMELREL. 01, CREATED)
 DT 01-NOV-1996 (TREMELREL. 01, LAST SEQUENCE UPDATE)
 DT 01-JAN-1999 (TREMELREL. 09, LAST ANNOTATION UPDATE)
 DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
 GN MSP1.
 OS PLASMODIUM FALCIPARUM.

OC EUKARYOTA; ALVEOLATA; APICOMPLEXA; HAEMOSPORIDA; PLASMODIUM.
 RN [1]

RP SEQUENCE FROM N.A.
 RA JONGWUTIWES S.; TANABE K.; KANBARA H.;
 RT "Sequence conservation in the C-terminal part of the precursor to the
 major merozoite surface proteins (MSP) of Plasmodium falciparum from
 field isolates."
 RL MOL. BIOCHEM. PARASITOL. 59:95-100(1993).
 DR EMBL; D13363; D1003128; -.
 DR PFAM; PF00008; EGF; 1.
 KW MEROZOITE; EGF-LIKE DOMAIN.

FT NON_TER 1
 SQ SEQUENCE 539 AA; 61046 MW; 398440E6 CRC32;

Query Match 54.8%; Score 57; DB 5; Length 539;
 Best Local Similarity 46.2%; Pred. No. 2.46e+00;
 Matches 6; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

Db 373 FKNNNFVGIADL 385
 QY 3 IKANSKFIGITEL 15

RESULT 13
 ID Q25966 PRELIMINARY; PRT; 539 AA.
 AC Q25966;
 DT 01-NOV-1996 (TREMELREL. 01, CREATED)
 DT 01-NOV-1996 (TREMELREL. 01, LAST SEQUENCE UPDATE)
 DT 01-JAN-1999 (TREMELREL. 09, LAST ANNOTATION UPDATE)
 DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
 GN MSP1.

OS PLASMODIUM FALCIPARUM.
 OC EUKARYOTA; ALVEOLATA; APICOMPLEXA; HAEMOSPORIDA; PLASMODIUM.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE; 93295445.
 RA JONGWUTIWES S.; TANABE K.; KANBARA H.;

```

RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSPl) of Plasmodium falciparum from
RT field isolates."
RL MOL. BIOCHEM. PARASITOL. 59:95-100(1993).
RL EMBL; D13357; D1003122; -.
DR PFAM; PF00008; EGF; 1.
KW MERZOITE; EGF-LIKE DOMAIN.
FT NON_TER 1
SQ SEQUENCE 539 AA; 61145 MW; DD8B38E CRC32;

Query Match      54.8%; Score 57; DB 5; Length 539;
Best Local Similarity 46.2%; Pred. No. 2.46e+00;
Matches 6; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

Db 373 FKNNFVGIADL 385
QY 3 IKANSKFIGITEL 15

RESULT 14
ID Q25973 PRELIMINARY; PRT; 539 AA.
AC Q25973;
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DI 01-JAN-1999 (TREMREL. 09, LAST ANNOTATION UPDATE)
DE MAJOR MERZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN MSPL.
OS PLASMODIUM FALCIPARUM.
OC EUKARYOTA; ALVEOLATA; APICOMPLEXA; HAEMOSPORIDA; PLASMODIUM.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93295445.
RA JONGWUTIWES S., TANABE K., KANBARA H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSPl) of Plasmodium falciparum from
RT field isolates."
RL MOL. BIOCHEM. PARASITOL. 59:95-100(1993).
DR EMBL; D13361; D1003126; -.
DR PFAM; PF00008; EGF; 1.
KW MERZOITE; EGF-LIKE DOMAIN.
FT NON_TER 1
SQ SEQUENCE 539 AA; 61047 MW; D7140867 CRC32;

Query Match      54.8%; Score 57; DB 5; Length 539;
Best Local Similarity 46.2%; Pred. No. 2.46e+00;
Matches 6; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

Db 373 FKNNFVGIADL 385
QY 3 IKANSKFIGITEL 15

RESULT 15
ID Q25971 PRELIMINARY; PRT; 539 AA.
AC Q25971;
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DI 01-JAN-1999 (TREMREL. 09, LAST ANNOTATION UPDATE)
DE MAJOR MERZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN MSPL.
OS PLASMODIUM FALCIPARUM.
OC EUKARYOTA; ALVEOLATA; APICOMPLEXA; HAEMOSPORIDA; PLASMODIUM.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93295445.
RA JONGWUTIWES S., TANABE K., KANBARA H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSPl) of Plasmodium falciparum from
RT field isolates."
RL MOL. BIOCHEM. PARASITOL. 59:95-100(1993).
DR EMBL; D13359; D1003124; -.
DR PFAM; PF00008; EGF; 1.
KW MERZOITE; EGF-LIKE DOMAIN.

```

```

FT NON_TER 1
SQ SEQUENCE 539 AA; 61017 MW; 7A8981F1 CRC32;

Query Match      54.8%; Score 57; DB 5; Length 539;
Best Local Similarity 46.2%; Pred. No. 2.46e+00;
Matches 6; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

Db 373 FKNNFVGIADL 385
QY 3 IKANSKFIGITEL 15

```

Search completed: Tue Aug 17 15:50:13 1999
Job time : 69 secs.

This Page Blank (uspto)

W P S R L
***** (TM)
Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Tue Aug 17 16:06:18 1999; MasPar time 3.99 Seconds
79.957 Million cell updates/sec
Tabular output not generated.

Title: >US-09-049-847-1
Description: (1-15) from US09049847.pep
Perfect Score: 104
Sequence: 1 QYIKANSKEFIGITEL 15
Scoring table: PAM 150
Gap 15

Searched: 170751 seqs, 21266608 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 18.668; Variance 53.652; scale 0.348

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					
Result No.	Score	Query Match	Length	Description	Pred. No.
1	104	100.0	15 38	T-cell epitope peptid	4.95e-05
2	104	100.0	15 23	Tetanus toxoid univer	4.95e-05
3	104	100.0	15 27	Universal T-cell epit	4.95e-05
4	104	100.0	15 1	Tetanus toxin epitope	4.95e-05
5	104	100.0	15 37	Tetanus toxoid epitop	4.95e-05
6	104	100.0	15 35	Universal helper T-ce	4.95e-05
7	104	100.0	15 36	Tetanus toxin fragmen	4.95e-05
8	104	100.0	16 27	T-cell stimulatory pe	4.95e-05
9	104	100.0	17 16	Tetanus toxin helper	4.95e-05
10	104	100.0	17 19	Tetanus toxin helper	4.95e-05
11	104	100.0	17 13	Helper T cell epitope	4.95e-05
12	104	100.0	17 17	T-cell antigen Tt2 pe	4.95e-05
13	104	100.0	27 16	IgE CH4 region contg.	4.95e-05
14	104	100.0	27 13	LHRH-containing immun	4.95e-05
15	104	100.0	29 16	IgE CH4 region contg.	4.95e-05
16	104	100.0	30 10	HIV antigen fragment.	4.95e-05

17	104	100.0	37 13	R65383	Universal immunostimu	4.95e-05
18	104	100.0	37 13	R65389	Universal immunostimu	4.95e-05
19	104	100.0	47 13	R62723	LHRH-containing immun	4.95e-05
20	104	100.0	50 20	W06131	Anti-cholesteryl este	4.95e-05
21	104	100.0	573 2	P70345	Portion of B fragment	4.95e-05
22	97	93.3	31 20	W06129	Anti-cholesteryl este	3.91e-04
23	97	93.3	63 3	R14263	Immunogenic branched	3.91e-04
24	96	92.3	14 8	R46509	Tetanus toxoid residu	5.23e-04
25	96	92.3	14 15	R78918	Tetanus toxoid 830-84	5.23e-04
26	96	92.3	14 20	W03003	Carrier peptide for a	5.23e-04
27	96	92.3	14 27	W35437	T-cell stimulatory pe	5.23e-04
28	96	92.3	14 15	R75943	T helper epitope from	5.23e-04
29	96	92.3	14 13	R70910	Tetanus toxoid 830-84	5.23e-04
30	96	92.3	14 30	W50108	Pan DR binding peptid	5.23e-04
31	96	92.3	14 1	R06309	Tetanus toxin epitope	5.23e-04
32	96	92.3	14 6	R33497	T helper peptide teta	5.23e-04
33	96	92.3	14 26	R74160	Antigenic peptide TT	5.23e-04
34	96	92.3	15 23	W11506	Tetanus toxoid mutant	5.23e-04
35	96	92.3	15 37	W73221	Tetanus toxoid epitop	5.23e-04
36	96	92.3	27 15	R78713	HBV specific cytotoxi	5.23e-04
37	96	92.3	27 6	R33507	T helper epitope/HSV	5.23e-04
38	96	92.3	29 33	W48992	Lipidated vaccine 2 a	5.23e-04
39	96	92.3	29 33	W48991	Lipidated vaccine 1 a	5.23e-04
40	96	92.3	30 23	W08413	Synthetic lipopeptide	5.23e-04
41	96	92.3	50 28	W46447	CETP B cell epitope/t	5.23e-04
42	96	92.3	50 20	W06132	Anti-cholesteryl este	5.23e-04
43	94	90.4	14 26	R74167	TT 830-843 based anti	9.39e-04
44	93	89.4	19 35	W78830	Tetanus toxoid protei	1.26e-03
45	91	87.5	14 1	R06318	Tetanus toxin epitope	2.24e-03

ALIGNMENTS

RESULT 1
ID W67578 standard; peptide; 15 AA.
AC W67578;
DT 02-MAR-1999 (first entry)
DE T-cell epitope peptide #4 for chimeric fimbriae/T-cell epitope peptide.
KW Chimeric: non-typable Haemophilus influenzae; fimbriae; T-cell epitope;
KW immunogenic composition; immune response.
OS Synthetic.
PN US5843464-A.
PD 01-DEC-1998.
PF 02-JUN-1995; 460502.
PR 02-JUN-1995; US-460502.
PA (OHIS) UNIV OHIO STATE.
PI Bakaletz LO, Kaumaya PTP;
DR WPI; 99-044514/04.
PT Synthetic chimeric fimbriae peptide - useful for vaccination against
PT non-typable Haemophilus influenzae
PS Disclosure: Column 4; 16pp; English.
CC The invention relates to the manufacture of a synthetic chimeric peptide
CC comprising a non-typable Haemophilus influenzae fimbriae peptide fused via
CC a linker peptide to a T-cell epitope peptide. The chimeric peptide is
CC used in immunogenic compositions which induce an immune response against
CC non-typable Haemophilus influenzae. This sequence represents an example
CC of a T-cell epitope peptide used to generate the chimeric peptide.
SQ Sequence 15 AA:
Query Match 100.0%; Score 104; DB 38; Length 15;
Best Local Similarity 100.0%; Pred. No. 4.95e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 1 qyikanskfigitel 15
QY 1 QYIKANSKEFIGITEL 15
RESULT 2
ID W11505 standard; Protein; 15 AA.
AC W11505;
DT 24-SEP-1997 (first entry)
DE Tetanus toxoid universal Th epitope IT830.

KW Humanised antibody; anti-Fc receptor; H22; bifunctional; bispecific;
 KW fusion protein; chimera; tetanus toxoid; helper T cell epitope;
 KW antigen presentation; ds.
 OS Clostridium tetani.
 PN WO9640789-A1.
 PD 19-DEC-1996.
 PF 07-JUN-1996; U09988.
 PR 07-JUN-1995; US-484172.
 PA (MEDA-) MEDAREX INC.
 PI Deo YM, Goldstein J, Graziano R, Somasundaram C;
 DR WPI; 97-052242/05.
 DT N-PSDB; T59127.
 PT Recombinant, multi-specific anti-Fc receptor antibody molecules
 PT also comprise an anti-target portion, used for the treatment of
 PT cancer, autoimmune disease and pathogenic infection
 PS Example 7; Fig 24; 115pp; English.
 CC Synthetic DNA coding for the wild-type universal Th epitope from
 CC tetanus toxoid, designated TR830, was fused to the 3'-end of DNA
 CC encoding heavy chain sequences from the humanised anti-Fc gamma RI
 CC monoclonal antibody H22. The resulting fusion protein was shown to
 CC be significantly more efficient in antigen presentation and T cell
 CC stimulation than the TR830 epitope alone. A similar fusion
 CC construct was prepared coding for a mutant, antagonistic form of the
 CC epitope (designated TR83S) fused to the anti-Fc gamma RI. The
 CC Fab22-TR83S is at least 100 times more effective than TR83S in
 CC inhibiting T cell activation.
 SQ Sequence 15 AA;

Query Match 100.0%; Score 104; DB 23; Length 15;
 Best Local Similarity 100.0%; Pred. No. 4.95e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 qyikanskfigitel 15
 QY 1 QYIKANSKFIGITEL 15

RESULT 3

ID W35506 standard; peptide; 15 AA.
 AC W35506;
 DT 22-APR-1998 (first entry)
 DE Universal T-cell epitope peptide SEQ ID NO:8.
 KW T-cell stimulatory peptide; immunogen; non-dendritic; carrier; tumour;
 KW scaffold; inhibition; metastasis; wound healing; solid phase.
 OS Unidentified.
 PN WO9738011-A1.
 PD 16-OCT-1997.
 PF 03-APR-1997; D00146.
 PR 03-APR-1996; DK-000398.
 PA (PEPR-) PEPRSEARCH AS.
 PI Heegaard PMH, Jakobsen PH;
 DR WPI; 97-512645/47.
 PT Non-dendritic peptide carrier linked to a solid phase - useful as a
 PT diagnostic agent and as a scaffold for production of chemical
 PT derivatives

PS Example 20; Page 124; 262pp; English.
 CC A non-dendritic peptide carrier (A) has been developed which is coupled
 CC through a linker to a solid phase, forming a complex of (A)-solid phase.
 CC Where (A) comprises 10-50 amino acids capable of forming a secondary
 CC structure in a benign buffer after liberation from the solid phase, and
 CC further the (A)-solid phase complex comprises an immunogenic substance
 CC and/or an immune mediator coupled on (A). The present sequence
 CC represents a peptide used in an example from the present invention. An
 CC (A)-solid phase complex can be used as a scaffold for the production of
 CC chemical derivatives, characterised by covalently attaching molecules at
 CC attachment points. Alternatively (A) is used as a scaffold-peptide for
 CC the incorporation into an immunostimulating complex (iscom) resulting in
 CC (A)-iscom complex which is used for the chemical coupling of antigenic
 CC substances in an aqueous solution by conjugation. (A) derivatised with
 CC one or more peptides having fibronectin-, laminin- or vitronectin-like
 CC binding activities can be used for the promotion of cell-attachment to
 CC plastic surfaces, in particular to inhibit tumour growth and metastasis,
 CC and for promotion of wound healing. Also a derivatised (A) can be used

CC for the selection of specifically-binding aptamers or as a diagnostic
 CC agent. Such diagnostic-(A) molecules could be used to detect molecules
 CC derived from or indicative of pregnancy or of a disease, such as an
 CC infectious, autoimmune or cancerous disease.
 SQ Sequence 15 AA;

Query Match 100.0%; Score 104; DB 27; Length 15;
 Best Local Similarity 100.0%; Pred. No. 4.95e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 qyikanskfigitel 15
 QY 1 QYIKANSKFIGITEL 15

RESULT 4

ID R06310 standard; protein; 15 AA.
 AC R06310;
 DT 04-DEC-1990 (first entry)
 DE Tetanus toxin epitope.
 KW Tetanus toxin; vaccine; major histocompatibility complex; MHC;
 KW antimalarial.
 OS Synthetic.
 PN EP-378881-A.
 PD 25-JUL-1990.
 PF 27-DEC-1989; 203318.
 PR 17-JAN-1989; IT-019110.
 PR 16-NOV-1989; IT-022409.
 PA (ENIE) ENIRICERCH SPA.
 PI Pessi A, Bianchi E, Verdini AS, Corradini G;
 DR WPI; 90-223582/30.
 PT Synthetic peptide(s) corresp. to tetanus toxin epitope(s) - used
 PT as universal carriers for prepn. of immunogenic conjugate(s) for
 PT use as vaccines.
 PS Claim 1; Page 17; 20pp; English.
 CC Epitopic peptides may be used with synthetic hapten derived from
 CC a pathogen to generate an immune response to the pathogen.
 CC Peptides are recognised by numerous T-helper cell clones within
 CC the context of a wide range of alleles of the human MHC.
 CC The peptides may be used in an antimalarial vaccine inducing Ab.
 CC response to P.falciparum.
 SQ Sequence 15 AA;

Query Match 100.0%; Score 104; DB 1; Length 15;
 Best Local Similarity 100.0%; Pred. No. 4.95e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 qyikanskfigitel 15
 QY 1 QYIKANSKFIGITEL 15

RESULT 5

ID W73220 standard; Protein; 15 AA.
 AC W73220;
 DT 25-JAN-1999 (first entry)
 DE Tetanus toxoid epitope.
 KW Multispecific single chain antibody; antibody H22; tumour cell; therapy;
 KW antibody-dependent cellular cytotoxicity; ADCC; HER 2/neu; infection;
 KW epidermal growth factor receptor; breast cancer; ovarian cancer.
 OS Synthetic.
 PN US5837243-A.
 PD 17-NOV-1998.
 PF 07-JUN-1996; 661052.
 PR 07-JUN-1996; US-661052.
 PR 07-JUN-1995; US-484172.
 PA (MEDA-) MEDAREX INC.
 PI Deo YM, Goldstein J, Graziano R, Somasundaram C;
 DR WPI; 99-023374/02.
 PT Specific killing of tumour cells - using a multi-specific molecule
 PT comprising an anti-Fc receptor antibody and a portion which binds to
 PT a target cell
 PS Example 7; Column 27; 57pp; English.

CC This sequence represents a tetanus toxoid epitope and is recognised
 CC by the multispecific single chain antibody designated H22. The
 CC antibody can be used in the method of the invention for inducing
 CC antibody-dependent cellular cytotoxicity (ADCC) against a tumour cell
 CC which is characterised by overexpression of HER 2/neu or epidermal growth
 CC factor receptor (EGFR), comprises contacting the tumour cell with a
 CC multispecific protein molecule (preferably a single chain antibody)
 CC comprising: (a) an anti-PC receptor antibody or an antigen binding
 CC fragment; (b) a portion which binds to HER 2/neu; and (c) a portion which
 CC binds to EGFR. The method can be used for treating cancers especially
 CC breast cancer or ovarian cancer. The multispecific antibody can also
 CC be administered prophylactically to vaccinate a subject against infection
 CC by a target cell.
 SQ Sequence 15 AA;

Query Match 100.0%; Score 104; DB 37; Length 15;
 Best Local Similarity 100.0%; Pred. No. 4.95e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 qvikanskfigitel 15
 QY 1 QYIKANSKFIGITEL 15

RESULT 6

ID W11321 standard; peptide; 15 AA.
 AC W11321:
 DT 26-NOV-1998 (first entry)
 DE Universal helper T-cell epitope P2 derived from tetanus toxin.
 KW Liver stage; Plasmodium; Navy Yoelli Liver Stage 3 antibody; NYLS3;
 KW hepatic and erythrocytic stage protein; PyHEP17; vaccine;
 KW malaria parasite; tetanus toxin; P2; helper T-cell epitope.
 OS Synthetic.
 OS Clostridium tetani.
 PN US3814617-A.
 PD 29-SEP-1998.
 PF 07-OCT-1994; 319704.
 PR 07-OCT-1994; US-319704.
 PA (USNA) US SEC OF NAVY.
 PI Charoenvit Y, Doolan DL, Hedstrom RC, Hoffman SL;
 DR WPI: 98-541794/46.
 PT Vaccine for protecting mammal against infection by malaria caused by
 PT Plasmodium species - comprises a first nucleic acid encoding a first
 PT polypeptide capable of eliciting an immune reaction against an
 PT antigen expressed during the liver
 PS Disclosure; Column 12; 24pp; English.
 CC W11321-22 represent universal helper T-cell epitopes derived from
 CC tetanus toxin. They are used to enhance host immune response to
 CC vaccines. The specification describes a Plasmodium yoelli liver stage
 CC 17 kDa hepatic and erythrocytic stage protein designated PyHEP17. This
 CC protein elicits a response from an Ig1 monoclonal antibody designated
 CC Navy Yoelli Liver Stage 3 (NYLS3). This antibody does not recognise
 CC sporozoites, but does recognise P. yoelli liver stage parasites. NYLS3
 CC eliminates upto 90% of liver stage parasites. The specification describes
 CC a vaccine for reducing the severity or incidence of infection by a
 CC malaria parasite of the genus Plasmodium. The DNA vaccine comprises
 CC exon 1 and part of exon 2 of the PyHEP17 gene.
 SQ Sequence 15 AA;

Query Match 100.0%; Score 104; DB 35; Length 15;
 Best Local Similarity 100.0%; Pred. No. 4.95e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 qvikanskfigitel 15
 QY 1 QYIKANSKFIGITEL 15

RESULT 7

ID W67033 standard; peptide; 15 AA.
 AC W67033;
 DT 15-DEC-1998 (first entry)
 DE Tetanus toxin fragment (residues 830-844).

KW Tetanus toxin; vaccine; antibody; carbohydrate peptide conjugate;
 KW dextrimeric poly-lysine; epitope; tumour.
 OS Clostridium tetani.
 PN W09843677-A1.
 PD 08-OCT-1998.
 PF 27-MAR-1998; E01922.
 PR 27-MAR-1997; US-041726.
 PA (INSP) INST PASIEUR.
 PI Ray S, Cantacuzene D, Leclerc C, Lo-man R;
 DR WPI: 98-557071/47.
 PT Carbohydrate peptide conjugate used as vaccine - comprises carrier
 PT with dextrimeric poly-lysine enabling multiple epitopes to be
 PT covalently attached
 PS Disclosure; Page 13; 55pp; English.
 CC The invention relates to a new carbohydrate peptide conjugate, which
 CC comprises a carrier with a dextrimeric poly-lysine enabling multiple
 CC epitopes to be covalently attached to it. Also claimed are: (1) an
 CC antibody purified from biological fluid or cells of organisms
 CC administered with the carbohydrate peptide conjugate, and (2) a diagnosis
 CC kit comprising antigen-specific antibodies elicited by immunisation with
 CC the carbohydrate peptide conjugate. The peptide conjugate, antibody and
 CC diagnosis kit are used to provide pharmaceutical compositions and
 CC vaccines against tumours. These can be used to support an immune response
 CC against viral infections caused by hepatitis virus, HIV or cytomegalo
 CC virus. They can be used to enhance immune responses, especially B- and T-
 CC cell responses, of humans and animals against bacterial infections. The
 CC carbohydrate peptide conjugate stimulates the antibody and T-cell
 CC response without stimulating undesired immune responses. The composition
 CC is capable of increasing the survival of tumour bearing humans and
 CC animals. The present sequence corresponds to residues 830-844 of tetanus
 CC toxin. The synthetic peptide corresponding to this sequence may be used
 CC as an epitope in a carbohydrate peptide conjugate.
 SQ Sequence 15 AA;

Query Match 100.0%; Score 104; DB 36; Length 15;
 Best Local Similarity 100.0%; Pred. No. 4.95e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 qvikanskfigitel 15
 QY 1 QYIKANSKFIGITEL 15

RESULT 8
 ID W35445 standard; peptide; 16 AA.
 AC W35445;
 DT 22-APR-1998 (first entry)
 DE T-cell stimulatory peptide SEQ ID NO:51.
 KW T-cell stimulatory peptide; immunogen; non-dendritic; carrier; tumour;
 KW scaffold; inhibition; metastasis; wound healing; solid phase.
 OS Unidentified.
 PN W09738011-A1.
 PD 16-OCT-1997.
 PF 03-APR-1997; D00146.
 PR 03-APR-1996; DK-000398.
 PA (PEPR) PEPRSEARCH AS.
 PI Heegaard PMH, Jakobsen PH;
 DR WPI: 97-512645/47.
 PT Non-dendritic peptide carrier linked to a solid phase - useful as a
 PT diagnostic agent and as a scaffold for production of chemical
 PT derivatives
 PS Claim 30; Page 199; 262pp; English.
 CC A non-dendritic peptide carrier (A) has been developed which is coupled
 CC through a linker to a solid phase, forming a complex of (A)-solid phase.
 CC Where (A) comprises 10-50 amino acids capable of forming a secondary
 CC structure in a benign buffer after liberation from the solid phase, and
 CC further the (A)-solid phase complex comprises an immunogenic substance
 CC and/or an immune mediator coupled on (A). The present sequence
 CC represents a specifically claimed T-cell stimulatory peptide from the
 CC present invention. An (A)-solid phase complex can be used as a scaffold
 CC for the production of chemical derivatives, characterised by covalently
 CC attaching molecules at attachment points. Alternatively (A) is used as
 CC a scaffold-peptide for the incorporation into an immunostimulating

CC Complex (Iscom) resulting an (A)-Iscom complex which is used for the
 CC chemical coupling of antigenic substances in an aqueous solution by
 CC conjugation. (A) derivatised with one or more peptides having
 CC fibronectin-, laminin- or vitronectin-like binding activities can be
 CC used for the promotion of cell-attachment to plastic surfaces, in
 CC particular to inhibit tumour growth and metastasis, and for promotion
 CC of wound healing. Also a derivatised (A) can be used for the selection
 CC of specifically-binding aptamers or as a diagnostic agent. Such
 CC diagnostic-(A) molecules could be used to detect molecules derived from
 CC or indicative of pregnancy or of a disease, such as an infectious,
 CC autoimmune or cancerous disease.
 SQ Sequence 16 AA;

Query Match 100.0%; Score 104; DB 27; Length 16;
 Best Local Similarity 100.0%; Pred. No. 4.95e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 1 QYKANSKFIGITEL 15
 QY 1 QYKANSKFIGITEL 15

RESULT 9
 ID R82573 standard; peptide; 17 AA.
 AC R82573;
 DI 13-JUN-1996 (first entry)
 DE Tetanus toxin helper T cell epitope, TMI.
 KW IgE; CH4; immunoglobulin; epsilon; immunogen; helper T cell; epitope;
 KW vaccine; allergy; antibody; constant heavy chain.
 OS Clostridium tetani.
 PN W09526365-A1.
 PD 05-OCT-1995.
 PF 24-MAR-1995; U03741.
 PR 28-MAR-1994; US-218461.
 PR 25-OCT-1994; US-328912.
 PA (UNBI-) UNITED BIOMEDICAL INC.
 PI Wang CY;
 DR WPI; 95-351297/45.
 PT Synthetic peptide-based immunogen contg. IgE CH4 peptide and helper
 PT T cell epitope - useful for eliciting antibody prodn. for allergy
 PI treatment

PS Claim 3; Page 59; 87pp; English.
 CC R82571-91 are helper T cell epitopes which can be used in the
 CC preparation of a peptide immunogen that is useful in vaccines for
 CC treating allergic reactions. In the immunogen an IgE CH4 peptide
 CC is attached C-terminally to a series of amino acids including a
 CC helper T cell epitope. The immunogen may also opt. contain a fatty
 CC acid or fatty acid derivative, an invasin domain or alpha-NH2. The
 CC immunogen produces high titres of antibodies to the effector site
 CC in human IgE heavy chain (the CH4 domain peptide) which inhibit mast
 CC cell activation and reduce allergen-induced IgE prodn. The immunogens
 CC may be used in either a radially branching multimeric form or a
 CC linearly arranged monomeric form.
 SQ Sequence 17 AA;

Query Match 100.0%; Score 104; DB 16; Length 17;
 Best Local Similarity 100.0%; Pred. No. 4.95e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 3 QYKANSKFIGITEL 17
 QY 1 QYKANSKFIGITEL 15

RESULT 10
 ID W05599 standard; peptide; 17 AA.
 AC W05599;
 DI 10-DEC-1996 (first entry)
 DE Tetanus toxin helper T cell epitope #1.
 KW Immunoglobulin; IgE; membrane protein; human; epsilon chain; hepatitis B;
 KW - membrane anchoring domain; helper T cell; surface antigen; core antigen;
 KW pertussis toxin; tetanus toxin; measles virus F protein; immunotherapy;
 KW Chlamydia trachomatis major outer membrane protein; immunogen; vaccine;

KW diphtheria toxin; plasmodium falciparum; circumsporozoite; E. coli TrAT;
 KW schistosoma mansoni; triose phosphate isomerase; allergic reaction;
 KW allergic rhinitis; food allergy; anaphylaxis; virally-induced asthma;
 KW antihistamine; decongestant; beta-2 agonist; immunosuppression;
 KW corticosteroid.
 OS Synthetic.
 PN W09612740-A1.
 PD 02-MAY-1996.
 PF 25-OCT-1995; U13841.
 PR 25-OCT-1994; US-328519.
 PA (UNBI-) UNITED BIOMEDICAL INC.
 PI Walfield AM, Wang CY;
 DR WPI; 96-230555/23.
 PT Peptide immunogen useful in treatment of allergy - comprises
 PT membrane-bound IgE epsilon-chain peptide synthesised linearly in
 PT tandem with T helper epitope peptide
 PS Claim 2; Page 18; 53pp; English.
 CC W05957-W05616 represent helper T cell epitopes used in the peptide
 CC immunogens of the invention. This sequence represents the tetanus toxin
 CC helper T cell antigen. The peptides of the invention contain one of
 CC these sequences, and a membrane-bound immunoglobulin E (IgE) fragment
 CC (see W05595 and W05596). The peptide immunogens of the invention can be
 CC used in vaccines for the immunotherapeutic treatment of allergic
 CC reactions, including allergic rhinitis, food allergies, anaphylaxis, or
 CC virally-induced asthma. The immunogens overcome the short effective
 CC period of antihistamines, decongestants, and beta-2 agonists, while
 CC preventing the broad immunosuppression of corticosteroids. The peptides
 CC do not have the potential side effects of restlessness or sedation
 CC (associated with antihistamines), associated increased morbidity in
 CC asthmatics (as seen with beta-2 agonists) and adverse hormonal activities
 CC (observed in corticosteroid users).
 SQ Sequence 17 AA;

Query Match 100.0%; Score 104; DB 19; Length 17;
 Best Local Similarity 100.0%; Pred. No. 4.95e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3 QYKANSKFIGITEL 17
 QY 1 QYKANSKFIGITEL 15

RESULT 11
 ID R62692 standard; peptide; 17 AA.
 AC R62692;
 DI 10-SEP-1995 (first entry)
 DE Helper T cell epitope for use in universal immune stimulator.
 KW Helper T cell epitope; universal immune stimulator; invasin; hapten;
 KW vaccine; tetanus toxin.
 OS Clostridium tetani.
 PN W09425060-A.
 PD 10-NOV-1994.
 PF 28-APR-1994; U04832.
 PR 27-APR-1993; US-057166.
 PR 14-APR-1994; US-229275.
 PA (LADD/) LADD A E.
 PA (WANG/) WANG C Y.
 PA (ZAMB/) ZAMB T.
 PI Ladd AE, Wang CY, Zamb T;
 DR WPI; 94-357910/44.
 PT Immunogenic luteinising hormone releasing hormone peptide(s) -
 PT that suppress LH/HR activity in males and females
 PS Claim 7; Page 25; 213pp; English.
 CC Synthetic immunogenic peptides are provided in which a universal immune
 CC stimulator is linked to a peptide or protein hapten containing B cell
 CC and/or cytotoxic T lymphocyte epitopes, giving a product which causes
 CC potent immune responses to the coupled peptide or protein. The
 CC stimulator consists of (A) a promiscuous helper T cell epitope (Th)
 CC which elicits an immune response to the coupled peptide in members of
 CC a heterogeneous population expressing diverse HLA phenotypes, and (B)
 CC an adjuvant peptide sequence from the invasin protein of Yersinia.
 CC Spacer amino acid sequences (e.g. Gly-Gly) can be provided between the
 CC invasin and Th domains and between the immune stimulator and hapten

CC components. When the hapten is LHRH, then optionally the invasive domain
 CC can be omitted from the immune stimulator component.
 CC The present sequence represents a tetanus toxin helper T cell epitope
 CC which can be used as Th in the immune stimulator.
 SQ Sequence 17 AA;

Query Match 100.0%; Score 104; DB 13; Length 17;
 Best Local Similarity 100.0%; Pred. No. 4.95e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3 qvikanskfigitel 17
 QY 1 QYIKANSKFIGITEL 15
 |||||

RESULT 12
 ID R88395 standard; Peptide; 17 AA.
 AC R88395;
 DT 12-JUN-1996 (first entry)
 DE T-cell antigen IT2 peptide.
 KW T-antigen; vaccine; antibody; T-cell; T-lymphocyte;
 KW alpha-helix; coiled-coil heterodimer; core peptide; subunit.
 OS Synthetic.
 PN W09531480-A1.
 PD 23-NOV-1995.
 PE 18-MAY-1995; CAC0293.
 PR 18-MAY-1994; US-245507.
 PA (SPIS-) SPI SYNTHETIC PEPTIDES INC.
 PI Cachia PJ, Hodges RS, Houston ME, Irvin RT, Kay CM;
 PI Zhou NE;
 DR WPI: 96-010880/01.
 PT Heterodimeric polypeptide immunogen in coiled-coil configuration
 PT with different antigens on each sub:unit - useful in vaccines and
 PT for antibody prodn.
 PS Claim 7; Page 61; 95pp; English.
 CC This T-cell antigen IT2 peptide may be attached to a core peptide
 CC contained in one of the 2 subunits of an alpha-helical coiled-coil
 CC heterodimer. Each core peptide is comprised of terminal and
 CC internal AA repeat sequences. This peptide antigen is attached
 CC to the core peptide through covalent linkages to certain AA of the
 CC internal repeats. The 2 subunits of the heterodimer are arranged
 CC in a stable alpha-helical coiled-coil configuration having a 1:1
 CC stoichiometry, and the peptide antigen is disposed toward the outer
 CC surfaces of the configuration. The heterodimer may be used as a
 CC synthetic vaccine (optionally multivalent) or to generate
 CC antibodies.
 SQ Sequence 17 AA;

Query Match 100.0%; Score 104; DB 17; Length 17;
 Best Local Similarity 100.0%; Pred. No. 4.95e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 qvikanskfigitel 15
 QY 1 QYIKANSKFIGITEL 15
 |||||

RESULT 13
 ID R82596 standard; peptide; 27 AA.
 AC R82596;
 DT 13-JUN-1996 (first entry)
 DE IgE CH4 region contg. peptide immunogen for treating allergies.
 KW IgE; CH4; immunoglobulin; epsilon; immunogen; helper T cell; epitope;
 KW vaccine; allergy; antibody; constant heavy chain.
 OS Synthetic.
 PN W09526365-A1.
 PD 05-OCT-1995.
 PE 24-MAR-1995; U03741.
 PR 28-MAR-1994; US-218461.
 PR 25-OCT-1994; US-328912.
 PA (UNBL-) UNITED BIOMEDICAL INC.
 PI Wang CY;
 DR WPI: 95-351297/45.

PT Synthetic peptide-based immunogen contg. IgE CH4 peptide and helper
 PT T cell epitope - useful for eliciting antibody prodn. for allergy
 PT treatment.
 PS Claim 5; Page 62; 87pp; English.
 CC R82592-R82600 and R83560-R83581 are peptide immunogens that are
 CC useful in vaccines for treating allergic reactions. In the immunogens,
 CC an IgE CH4 peptide is attached C-terminally to a series of amino acids
 CC including a helper T cell epitope. The immunogen may also opt. contain
 CC a fatty acid or fatty acid derivative, an invasive domain or alpha-NH2.
 CC The immunogen produces high titres of antibodies to the effector site
 CC in human IgE heavy chain (the CH4 domain peptide) which inhibit mast
 CC cell activation and reduce allergen-induced IgE prodn. The immunogens
 CC may be used in either a radially branching multimeric form or a
 CC linearly arranged monomeric form.
 SQ Sequence 27 AA;

Query Match 100.0%; Score 104; DB 16; Length 27;
 Best Local Similarity 100.0%; Pred. No. 4.95e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3 qvikanskfigitel 17
 QY 1 QYIKANSKFIGITEL 15
 |||||

RESULT 14
 ID R62701 standard; peptide; 27 AA.
 AC R62701;
 DT 10-SEP-1995 (first entry)
 DE LHRH-containing immunogenic peptide.
 KW Helper T cell epitope; universal immune stimulator; invasive; hapten;
 KW vaccine; LHRH; luteinising hormone releasing hormone; prostate;
 KW androgen-dependent carcinoma; antitumour; infertility;
 KW tetanus toxin.
 OS Synthetic.
 FH Key
 FT domain
 FT domain
 FT domain
 FT domain
 PN W09425060-A.
 PD 10-NOV-1994.
 PF 28-APR-1994; U04832.
 PR 27-APR-1993; US-057166.
 PR 14-APR-1994; US-229275.
 PA (LADD/) LADD A E.
 PA (WANG/) WANG C Y.
 PA (ZAMB/) ZAMB T.
 PA Ladd AE, Wang CY, Zamb T;
 DR WPI: 94-357910/44.
 PT Immunogenic luteinising hormone releasing hormone peptide(s) -
 PT that suppress LHRH activity in males and females
 PS Claims 8, 12; Page 84; 213pp; English.
 CC Synthetic immunogenic peptides are provided in which a universal immune
 CC stimulator is linked to a peptide or protein haptens containing B cell
 CC and/or cytotoxic T lymphocyte epitopes, giving a product which causes
 CC potent immune responses to the coupled peptide or protein. The
 CC stimulator consists of (A) a promiscuous helper T cell epitope (Th)
 CC which elicits an immune response to the coupled peptide in members of
 CC a heterogeneous population expressing diverse HLA phenotypes, and (B)
 CC an adjuvant peptide sequence from the invasive protein of Yersinia.
 CC Spacer amino acid sequences (e.g. Gly-Gly) can be provided between the
 CC invasive and Th domains and between the immune stimulator and haptens
 CC components. When the haptens is LHRH, then optionally the invasive domain
 CC can be omitted from the immune stimulator component.
 CC The present sequence represents an LHRH-containing, invasive-free
 CC immunogenic peptide as above which can be used as a potent vaccine for
 CC treating e.g. prostatic hyperplasia, androgen-dependent carcinoma,
 CC prostatic carcinoma, testicular carcinoma, endometriosis, benign
 CC uterine tumours, recurrent functional ovarian cysts, (severe)
 CC premenstrual syndrome or oestrogen-dependent breast cancer, or for
 CC induction of infertility.
 CC This sequence is particularly preferred.

```

SQ Sequence 27 AA;
Query Match 100.0%; Score 104; DB 13; Length 27;
Best Local Similarity 100.0%; Pred. No. 4.95e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3 qvikanskfigitel 17
   |||||
QY 1 QYIKANSKFIGITEL 15

RESULT 15
ID R83561 standard; peptide; 29 AA.
AC R83561;
DT 13-JUN-1996 (first entry)
DE IgE CH4 region contg. peptide immunogen for treating allergies.
KW IgE; CH4; immunoglobulin; epsilon; immunogen; helper T cell; epitope;
KW vaccine; allergy; antibody; constant heavy chain.
OS Synthetic.
PN WO9526365-A1.
PR 05-OCT-1995.
PF 24-MAR-1995; U03741.
PR 28-MAR-1994; US-218461.
PR 25-OCT-1994; US-328912.
PA (UNBI-) UNITED BIOMEDICAL INC.
PI Wang CY;
DR WPI; 95-351297/45.
PT Synthetic peptide-based immunogen contg. IgE CH4 peptide and helper
PT T cell epitope - useful for eliciting antibody prodn. for allergy
PT treatment
PS Claim 5; Page 68-69; 87pp; English.
CC R82592-R82600 and R83560-R83581 are peptide immunogens that are
CC useful in vaccines for treating allergic reactions. In the immunogens,
CC an IgE CH4 peptide is attached C-terminally to a series of amino acids
CC including a helper T cell epitope. The immunogen may also opt. contain
CC a fatty acid or fatty acid derivative, an invasin domain or alpha-NH2.
CC The immunogen produces high titres of antibodies to the effector site
CC in human IgE heavy chain (the CH4 domain peptide) which inhibit mast
CC cell activation and reduce allergen-induced IgE prodn. The immunogens
CC may be used in either a radially branching multimeric form or a
CC linearly arranged monomeric form.
SQ Sequence 29 AA;

Query Match 100.0%; Score 104; DB 16; Length 29;
Best Local Similarity 100.0%; Pred. No. 4.95e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3 qvikanskfigitel 17
   |||||
QY 1 QYIKANSKFIGITEL 15

RESULT 16
ID R44398 standard; peptide; 30 AA.
AC R44398;
DT 08-NOV-1994 (first entry)
DE HIV antigen fragment.
KW HIV; human immunodeficiency virus; immunisation; monoclonal antibody.
OS Human immunodeficiency virus.
PN TW-208717-A.
PR 01-JUL-1993.
PR 24-APR-1992; 103240.
PR 24-APR-1992; TW-103240.
PA (CHIN/) CHIN L.
PI Chin L;
DR WPI; 93-335491/42.
PT Induction of neutralising human monoclonal antibodies against
PT human immuno: deficiencies - by sepg. peripheral mono:nuclear cells
PT from blood using density gradient centrifugation, and treating
PT cells by L-leucyl-L-leucine methyl ester etc.
TS Claim 1; ; 36pp; Taiwanese.
CC The invention relates to a method of assessing human
CC immunodeficiency virus and producing human immunodeficiency

```

```

CC antibodies by in-vitro immunisation, which comprises: (a) separating
CC peripheral mononuclear cells from blood using density gradient
CC centrifugation; (b) treating the mononuclear cells with L-leucyl-L-
CC leucine methyl ester; and (c) using the present antigen fragment,
CC which is formed by coupled T and B cells, in a culture medium of
CC human serum, IL-2 and T cells to effect cultivation and achieve in
CC vitro immunisation.
SQ Sequence 30 AA;

Query Match 100.0%; Score 104; DB 10; Length 30;
Best Local Similarity 100.0%; Pred. No. 4.95e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 qvikanskfigitel 15
   |||||
QY 1 QYIKANSKFIGITEL 15

RESULT 17
ID R65383 standard; peptide; 37 AA.
AC R65383;
DT 21-SEP-1995 (first entry)
DE Universal immunostimulator having GG spacers.
KW Helper T cell epitope; universal immune stimulator; invasin; haptin;
KW tetanus toxin.
OS Synthetic.
PH Key Location/Qualifiers
FT domain 1..16 /note= "invasin domain"
FT domain 19..35 /note= "tetanus toxin helper T cell epitope"
PN WO9425060-A.
PD 10-NOV-1994.
PE 28-APR-1994; U04832.
PR 27-APR-1993; US-057166.
PR 14-APR-1994; US-229275.
PA (LADD/) LADD A E.
PA (WANG/) WANG C Y.
PA (ZAMB/) ZAMB T.
PI Ladd AE, Wang CY, Zamb I;
PI WPI; 94-357910/44.
PT Immunogenic luteinising hormone releasing hormone peptide(s) -
PT that suppress LHSH activity in males and females
PS Disclosure; Page 95; 213pp; English.
CC Synthetic immunogenic peptides are provided in which a universal immune
CC stimulator is linked to a peptide or protein haptin containing B cell
CC and/or cytotoxic T lymphocyte epitopes, giving a product which causes
CC potent immune responses to the coupled peptide or protein. The
CC stimulator consists of (A) a promiscuous helper T cell epitope (Th)
CC which elicits an immune response to the coupled peptide in members of
CC a heterogeneous population expressing diverse HLA phenotypes, and (B)
CC an adjuvant peptide sequence from the invasin protein of Yersinia.
CC Spacer amino acid sequences (e.g. Gly-Gly) can be provided between the
CC invasin and Th domains and between the immune stimulator and haptin
CC components. When the haptin is LHRH, then optionally the invasin domain
CC can be omitted from the immune stimulator component.
CC The present sequence is an example of an invasin-GG-Th-GG- immune
CC stimulator to which a haptin can be bonded.
SQ Sequence 37 AA;

Query Match 100.0%; Score 104; DB 13; Length 37;
Best Local Similarity 100.0%; Pred. No. 4.95e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 21 qvikanskfigitel 35
   |||||
QY 1 QYIKANSKFIGITEL 15

RESULT 18
ID R65389 standard; peptide; 37 AA.
AC R65389;
DT 21-SEP-1995 (first entry)

```

DE Universal immunostimulator having GG spacers.
 KW Helper T cell epitope; universal immune stimulator; invasin; haptens;
 KW tetanus toxin.
 OS Synthetic.
 FH Key
 FT domain 3..19 Location/Qualifiers
 FT /note= "tetanus toxin helper T cell epitope"
 FT 22..37
 FT /note= "invasin domain"
 FN W09425060-A.
 PD 10-NOV-1994.
 PE 28-APR-1994; U04832.
 PR 27-APR-1993; US-057166.
 PR 14-APR-1994; US-229275.
 PA (LADD/) LADD A E.
 PA (WANG/) WANG C Y.
 PA (ZAMB/) ZAMB T.
 PI Ladd AE, Wang CY, Zamb T;
 DR WPI: 94-357910/44.
 PT Immunogenic luteinising hormone releasing hormone peptide(s) -
 PI that suppress LHRH activity in males and females
 PS Disclosure; Page 95; 213pp; English.
 CC Synthetic immunogenic peptides are provided in which a universal immune
 CC stimulator is linked to a peptide or protein haptens containing B cell
 CC and/or cytotoxic T lymphocyte epitopes, giving a product which causes
 CC potent immune responses to the coupled peptide or protein. The
 CC stimulator consists of (A) a promiscuous helper T cell epitope (Th)
 CC which elicits an immune response to the coupled peptide in members of
 CC a heterogeneous population expressing diverse HLA phenotypes, and (B)
 CC an adjuvant peptide sequence from the invasin protein of fersinia.
 CC Spacer amino acid sequences (e.g. Gly-Gly) can be provided between the
 CC invasin and Th domains and between the immune stimulator and haptens
 CC components. When the haptens is LHRH, then optionally the invasin domain
 CC can be omitted from the immune stimulator component.
 CC The present sequence represents an LHRH-containing immunogenic peptide
 CC as above which can be used as a potent vaccine for treating e.g.
 CC prostatic hyperplasia, androgen-dependent carcinoma, prostatic
 CC carcinoma, testicular carcinoma, endometriosis, benign uterine tumours,
 CC recurrent functional ovarian cysts, (severe) premenstrual syndrome or
 CC oestrogen-dependent breast cancer, or for induction of infertility.
 SQ Sequence 37 AA;
 Query Match 100.0%; Score 104; DB 13; Length 37;
 Best Local Similarity 100.0%; Pred. No. 4.95e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 5 qvikansfigitel 19
 QY 1 QYIKANSKFIGITEL 15
 RESULT 19
 ID R62723 standard; peptide; 47 AA.
 AC R62723;
 DI 17-SEP-1995 (first entry)
 DE LHRH-containing immunogenic peptide.
 KW Helper T cell epitope; universal immune stimulator; invasin; haptens;
 KW vaccine; LHRH; luteinising hormone releasing hormone; prostate;
 KW androgen-dependent carcinoma; antitumour; infertility;
 KW tetanus toxin.
 OS Synthetic.
 FH Key
 FT domain 1..16 Location/Qualifiers
 FT /note= "invasin domain"
 FT 19..35
 FT /note= "tetanus toxin helper T cell epitope"
 FT 38..47
 FT /note= "LHRH haptens"
 FN W09425060-A.
 PD 10-NOV-1994.
 PE 28-APR-1994; U04832.
 PR 27-APR-1993; US-057166.
 PR 14-APR-1994; US-229275.
 PA (LADD/) LADD A E.
 PA (WANG/) WANG C Y.
 PA (ZAMB/) ZAMB T.
 PI Ladd AE, Wang CY, Zamb T;

DR WPI: 94-357910/44.
 PT Immunogenic luteinising hormone releasing hormone peptide(s) -
 PT that suppress LHRH activity in males and females
 PS Claim 8; Page 88; 213pp; English.
 CC Synthetic immunogenic peptides are provided in which a universal immune
 CC stimulator is linked to a peptide or protein haptens containing B cell
 CC and/or cytotoxic T lymphocyte epitopes, giving a product which causes
 CC potent immune responses to the coupled peptide or protein. The
 CC stimulator consists of (A) a promiscuous helper T cell epitope (Th)
 CC which elicits an immune response to the coupled peptide in members of
 CC a heterogeneous population expressing diverse HLA phenotypes, and (B)
 CC an adjuvant peptide sequence from the invasin protein of fersinia.
 CC Spacer amino acid sequences (e.g. Gly-Gly) can be provided between the
 CC invasin and Th domains and between the immune stimulator and haptens
 CC components. When the haptens is LHRH, then optionally the invasin domain
 CC can be omitted from the immune stimulator component.
 CC The present sequence represents an LHRH-containing immunogenic peptide
 CC as above which can be used as a potent vaccine for treating e.g.
 CC prostatic hyperplasia, androgen-dependent carcinoma, prostatic
 CC carcinoma, testicular carcinoma, endometriosis, benign uterine tumours,
 CC recurrent functional ovarian cysts, (severe) premenstrual syndrome or
 CC oestrogen-dependent breast cancer, or for induction of infertility.
 SQ Sequence 47 AA;
 Query Match 100.0%; Score 104; DB 13; Length 47;
 Best Local Similarity 100.0%; Pred. No. 4.95e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 21 qvikansfigitel 35
 QY 1 QYIKANSKFIGITEL 15
 RESULT 20
 ID W06131 standard; Peptide: 50 AA.
 AC W06131;
 DI 07-FEB-1997 (first entry)
 DE Anti-cholesterol ester transfer multivalent vaccine peptide.
 KW Cholesteryl ester transfer protein; CERP; antigen; vaccine;
 KW cardiovascular disease; atherosclerosis.
 OS Synthetic.
 FH Key
 FT misc_difference 1 Location/Qualifiers
 FT /note= "C-terminal Cys residue is present for use
 FT in linking the peptide to itself or other
 FT molecules"
 FT 2..15
 FT /label= T-cell epitope
 FT /note= "T-cell epitope comprises amino acids
 FT 830-843 of tetanus toxoid protein"
 FT 16..34
 FT /label= B-cell epitope
 FT /note= "B-cell epitope comprises amino acids
 FT 349-367 of human CERP"
 FT 35..50
 FT /label= B-cell epitope
 FT /note= "B-cell epitope comprises the C-terminal 16
 FT amino acids of human CERP, involved in
 FT neutral lipid binding or transfer activity"
 FN W09634888-A1.
 PD 07-NOV-1996.
 PE 01-MAY-1996; U06147.
 PR 01-MAY-1995; US-432483.
 PA (TCEL-) T CELL SCI INC.
 PI Rittershaus CW, Thomas LJ;
 DR WPI: 96-506103/50.
 PT Cholesteryl ester transfer protein B cell epitope linked to T cell
 PT epitope - used to generate vaccine to regulate CERP activity for
 PT decreasing the risk of developing a cardiovascular disease e.g.
 PT atherosclerosis
 PT Disclosure; Page 7; 72pp; English.
 PS A multivalent vaccine comprises an immunogenic helper T-cell
 CC epitope of tetanus toxoid protein covalently linked to the B-cell

Tue Aug 17 16:08:42 1999

CC epitopes of human cholesteryl ester transfer protein (CETP) (see
 CC also W06127). The vaccine elicits an immune response against
 CC endogenous CETP activity, and is used to treat or prevent a
 CC cardiovascular disease, such as atherosclerosis.

SQ Sequence 50 AA;
 Query Match 100.0%; Score 104; DB 20; Length 50;
 Best Local Similarity 100.0%; Pred. No. 4.95e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2 qvikanskfigitel 16
 QY 1 QYIKANSKFIGITEL 15

RESULT 21
 ID P70345 standard; Protein; 573 AA.
 AC P70345;
 DT 22-APR-1991 (first entry)
 DE Portion of B fragment and all of the C fragment of tetanus toxin.
 KW TT; vaccine.
 OS Clostridium tetani.
 PN EP-209281-A.
 PD 21-JAN-1987.
 PE 27-JUN-1986; 305029.
 PR 28-JUN-1985; GB-016442.
 PA (WEL) Wellcome Foundation Ltd.
 PI Fairweather NF;
 DR WPI; 87-015993/03.
 DR N-PSDB; N70545.
 PT Cloned DNA sequence coding for tetanus toxin - or its fragments
 PT contg. epitope used to express antigens for vaccine production.
 PS Clam 4; Fig 1; 36pp; English.
 CC Gene product comprises a tetanus toxin fragment, which may be
 CC expressed in a transformed host, and used as an antigen in vaccine
 CC production, against the disease.
 SQ Sequence 573 AA;

Query Match 100.0%; Score 104; DB 2; Length 573;
 Best Local Similarity 100.0%; Pred. No. 4.95e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 88 qvikanskfigitel 102
 QY 1 QYIKANSKFIGITEL 15

Search completed: Tue Aug 17 16:06:47 1999
 Job time : 29 secs.

M P E R L H (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Tue Aug 17 16:07:31 1999; MasPar time 7.53 Seconds
Tabular output not generated.
100.464 Million cell updates/sec

Title: >US-09-049-847-1
Description: (1-15) from US09049847.pep
Perfect Score: 104
Sequence: 1 QYIKANSKFIGITEL 15

Scoring table: PAM 150
Gap 15

Searched: 411786 seqs, 50406085 residues

Post-processing: Minimum Match 100%
Listing first 1000 summaries
Maximum DB seq length 15

Database: a-pending
1:P9 2:U60 3:U7 4:U80 5:U81 6:U82 7:U83 8:U84A 9:U84B
10:U85 11:U86 12:U87 13:U88 14:U89 15:U90 16:U91 17:U92
18:U93 19:NEWP 20:NEWU6 21:NEWU8 22:NEWU9

Statistics: Mean 19.566; Variance 56.870; scale 0.344

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match Length DB ID	Description Pred. No.
1	104	100.0	15 8 US-08-406- Sequence 2, Applicatio 2.56e-04
2	104	100.0	15 1 PCT-US99-0 Sequence 2, Applicatio 2.56e-04
3	104	100.0	15 5 US-08-161- Sequence 69, Applicati 2.56e-04
4	104	100.0	15 12 US-08-716- Sequence 4, Applicatio 2.56e-04
5	104	100.0	15 15 US-09-049- Sequence 1, Applicatio 2.56e-04
6	104	100.0	15 15 US-09-046- Sequence 2, Applicatio 2.56e-04
7	104	100.0	15 5 US-08-161- Sequence 69, Applicati 2.56e-04

Note: Post-processor removed 993 summaries from list due to search parameters chosen.

ALIGNMENTS

RESULT 1
ID US-08-406-916B-2 STANDARD; PRT; 15 AA.
XX
AC xxxxxx
XX
XX
DT
XX

DE Sequence 2, Application US/08406916B
XX
CC Sequence 2, Application US/08406916B
CC GENERAL INFORMATION:
CC APPLICANT: STEVENS, VERNON C.
CC TITLE OF INVENTION: VACCINES AND ANTIGENIC CONJUGATES
CC NUMBER OF SEQUENCES: 31
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MILLARD, SIDNEY W.
CC STREET: 7632 SLATE RIDGE BOULEVARD
CC CITY: REYNOLDSBURG
CC STATE: OHIO
CC COUNTRY: USA
CC ZIP: 43068
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB
CC COMPUTER: IBM PS/2 COMPATIBLE 486 DX 33
CC OPERATING SYSTEM: MS DOS 6.2
CC SOFTWARE: WORDPERFECT FOR MS DOS 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/406,916B
CC FILING DATE: 27 MAR 1995
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US92/08370
CC FILING DATE: 30 SEP 1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: KREMBLAS, FRANCIS T., JR.
CC REGISTRATION NUMBER: 22,773
CC REFERENCE/DOCKET NUMBER: URF 2 056 3 3
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (614) 575 2100
CC TELEFAX: (614) 575 2149
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 15 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE:
CC DESCRIPTION: peptide
CC HYPOTHETICAL: NO
CC FRAGMENT TYPE: 830-844 fragment of tetanus toxoid
CC FEATURE:
CC OTHER INFORMATION: T cell epitope
CC SEQUENCE 15 AA; 1725 MW; 1271 CN;
SQ

Query Match 100.0%; Score 104; DB 8; Length 15;
Best Local Similarity 100.0%; Pred. No. 2.56e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 QYIKANSKFIGITEL 15
QY 1 QYIKANSKFIGITEL 15

RESULT 2
ID PCT-US99-06325-2 STANDARD; PRT; 15 AA.
XX
AC xxxxxx
XX
DT
XX
XX
DE Sequence 2, Application PC/TUS9906325
XX
CC Sequence 2, Application PC/TUS9906325
CC GENERAL INFORMATION:
CC APPLICANT: Sudhir Paul
CC APPLICANT: Larry J. Smith
CC APPLICANT: Gennady Gololobov
CC TITLE OF INVENTION: Methods for Identifying Inducers and
CC TITLE OF INVENTION: Inhibitors of Catalytic Antibodies, Compositions and Their
CC TITLE OF INVENTION: Use
CC FILE REFERENCE: UNMC 63123
CC CURRENT APPLICATION NUMBER: PCT/US99/06325

CC CURRENT FILING DATE: 1999-03-23
CC EARLIER APPLICATION NUMBER: US 09/046,373
CC EARLIER FILING DATE: 1998-03-23
CC NUMBER OF SEQ ID NOS: 10
CC SOFTWARE: FastSeq for Windows Version 3.0
CC SEQ ID NO 2
CC LENGTH: 15
CC TYPE: PRT
CC ORGANISM: Clostridium tetani
CC SEQUENCE 15 AA; 1725 MW; 1271 CN;

Query Match 100.0%; Score 104; DB 1; Length 15;
Best Local Similarity 100.0%; Pred. No. 2.56e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 QYIKANSKFIGITEL 15
QY 1 QYIKANSKFIGITEL 15

RESULT 3

ID US-09-161-889-69 STANDARD; PRT; 15 AA.
AC xxxxxx
XX
DT
DE
XX

Sequence 69, Application US/08161889

Sequence 69, Application US/08161889

GENERAL INFORMATION:
APPLICANT: Geysen, H. Mario
APPLICANT: Rodda, Stuart J.
TITLE OF INVENTION: T-Cell Epitopes
NUMBER OF SEQUENCES: 75
CORRESPONDENCE ADDRESS:
ADDRESSEE: Grant D. Green
STREET: 4560 Horton St.
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/161,889
FILING DATE: 02-DEC-1993
CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/984,852
FILING DATE: 02-DEC-1992

ATTORNEY/AGENT INFORMATION:
NAME: Green, Grant D.
REGISTRATION NUMBER: 31,259
REFERENCE/DOCKET NUMBER: 0222.002

TELEPHONE: 510-601-2706
TELEFAX: 510-855-3542

INFORMATION FOR SEQ ID NO: 69:

SEQUENCE CHARACTERISTICS:

LENGTH: 15 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

SEQUENCE 15 AA; 1725 MW; 1271 CN;

Query Match 100.0%; Score 104; DB 5; Length 15;

Best Local Similarity 100.0%; Pred. No. 2.56e-04;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 QYIKANSKFIGITEL 15

QY 1 QYIKANSKFIGITEL 15

RESULT 5

ID US-09-049-847-1 STANDARD; PRT; 15 AA.

AC xxxxxx

XX

Db 1 QYIKANSKFIGITEL 15
QY 1 QYIKANSKFIGITEL 15

RESULT 4

ID US-08-716-249-4 STANDARD; PRT; 15 AA.

AC xxxxxx

XX

DT

DE

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

```
DT XX
DE XX
XX
Sequence 1, Application US/09049847
Sequence 1, Application US/09049847
GENERAL INFORMATION:
APPLICANT: Bay, Sylvie
APPLICANT: Cantacuzene, Daniele
APPLICANT: Leclerc, Claude
APPLICANT: Lo-Man, Richard
TITLE OF INVENTION: Multiple antigen alycopeptide carbohydrate, vaccine
FILE REFERENCE: 102.166A
CURRENT APPLICATION NUMBER: US/09/049,847
CURRENT FILING DATE: 1998-03-27
EARLIER APPLICATION NUMBER: 60/041,726
EARLIER FILING DATE: 1997-03-27
NUMBER OF SEQ ID NOS: 4
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 15
TYPE: PRT
ORGANISM: Clostridium tetani
SQ SEQUENCE 15 AA; 1725 MW; 1271 CN;

Query Match 100.0%; Score 104; DB 15; Length 15;
Best Local Similarity 100.0%; Pred. No. 2.56e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 QYIKANSKFIGITEL 15
QY 1 QYIKANSKFIGITEL 15

RESULT 6
ID US-09-046-373-2 STANDARD; PRT; 15 AA.
XX
AC xxxxxx
XX
DT
DE
XX
Sequence 2, Application US/09046373
Sequence 2, Application US/09046373
GENERAL INFORMATION:
APPLICANT: Sudhir Paul
APPLICANT: Larry J. Smith
APPLICANT: Gennady Gololobov
TITLE OF INVENTION: Methods for Identifying Inducers and
TITLE OF INVENTION: Inhibitors of Catalytic Antibodies, Compositions and Their
TITLE OF INVENTION: Use
FILE REFERENCE: UNMC 63123
CURRENT APPLICATION NUMBER: US/09/046,373
CURRENT FILING DATE: 1998-03-23
NUMBER OF SEQ ID NOS: 10
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 15
TYPE: PRT
ORGANISM: Clostridium tetani
SQ SEQUENCE 15 AA; 1725 MW; 1271 CN;

Query Match 100.0%; Score 104; DB 15; Length 15;
Best Local Similarity 100.0%; Pred. No. 2.56e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 QYIKANSKFIGITEL 15
QY 1 QYIKANSKFIGITEL 15

RESULT 7
ID US-08-161-889A-69 STANDARD; PRT; 15 AA.
```

```
XX
AC
XX
DT
XX
DE
XX
Sequence 69, Application US/08161889A
Sequence 69, Application US/08161889A
GENERAL INFORMATION:
APPLICANT: Geysen, H. Mario
APPLICANT: Rodda, Stuart J.
TITLE OF INVENTION: T-Cell Epitopes
NUMBER OF SEQUENCES: 74
CORRESPONDENCE ADDRESS:
ADDRESSEE: Grant D. Green
STREET: 4560 Horton St.
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/161,889A
FILING DATE: 02-DEC-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/984,852
FILING DATE: 02-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Green, Grant D.
REGISTRATION NUMBER: 31,259
REFERENCE/DOCKET NUMBER: 0222.002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 510-601-2706
TELEFAX: 510-655-3542
INFORMATION FOR SEQ ID NO: 69:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SQ SEQUENCE 15 AA; 1725 MW; 1271 CN;

Query Match 100.0%; Score 104; DB 5; Length 15;
Best Local Similarity 100.0%; Pred. No. 2.56e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 QYIKANSKFIGITEL 15
QY 1 QYIKANSKFIGITEL 15

Search completed: Tue Aug 17 16:10:08 1999
Job time : 157 secs.
```

This Page Blank (uspto)

W P S R L H

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Aug 17 15:07:04 1999; MasPar time 1.99 Seconds

Tabular output not generated. 76.451 Million cell updates/sec

Title: >US-09-049-847-1
Description: (1-15) from US09049847.pep
Perfect Score: 104
Sequence: 1 QYIKANSKFIGITEL 15

Scoring table: PAM 150
Gap 15

Searched: 106580 seqs, 10152877 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:PCT9_COMB 4:backfiles1

Statistics: Mean 17.539; Variance 50.688; scale 0.346

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	104	100.0	15	3	PCT-US93-1 Sequence 69, Applicati	1.64e-05
2	104	100.0	15	2	US-08-460- Sequence 7, Applicatio	1.64e-05
3	104	100.0	15	2	US-08-661- Sequence 6, Applicatio	1.64e-05
4	104	100.0	15	2	US-08-319- Sequence 10, Applicati	1.64e-05
5	104	100.0	17	3	PCT-US95-1 Sequence 7, Applicatio	1.64e-05
6	104	100.0	17	3	PCT-US95-0 Sequence 23, Applicati	1.64e-05
7	104	100.0	17	2	US-08-488- Sequence 4, Applicatio	1.64e-05
8	104	100.0	17	2	US-08-446- Sequence 13, Applicati	1.64e-05
9	104	100.0	27	2	US-08-488- Sequence 13, Applicati	1.64e-05
10	104	100.0	31	3	PCT-US93-1 Sequence 63, Applicati	1.64e-05
11	104	100.0	37	2	US-08-446- Sequence 53, Applicati	1.64e-05
12	104	100.0	37	2	US-08-488- Sequence 57, Applicati	1.64e-05
13	104	100.0	37	2	US-08-488- Sequence 53, Applicati	1.64e-05
14	104	100.0	37	2	US-08-446- Sequence 35, Applicati	1.64e-05
15	104	100.0	37	2	US-08-488- Sequence 63, Applicati	1.64e-05
16	104	100.0	47	2	US-08-488- Sequence 35, Applicati	1.64e-05
17	104	100.0	47	2	US-08-446- Sequence 35, Applicati	1.64e-05
18	97	93.3	24	3	PCT-US95-0 Sequence 110, Applicat	1.31e-04
19	97	93.3	24	3	PCT-US92-0 Sequence 25, Applicati	1.31e-04
20	97	93.3	27	3	PCT-US92-0 Sequence 32, Applicati	1.31e-04
21	96	92.3	14	3	PCT-US95-0 Sequence 95, Applicati	1.76e-04
22	96	92.3	14	3	PCT-US92-0 Sequence 15, Applicati	1.76e-04
23	96	92.3	14	3	PCT-US92-0 Sequence 30, Applicati	1.76e-04

24	96	92.3	14	1	US-08-186- Sequence 5, Applicatio	1.76e-04
25	96	92.3	14	2	US-08-465- Sequence 18, Applicati	1.76e-04
26	96	92.3	14	1	US-08-305- Sequence 5, Applicatio	1.76e-04
27	96	92.3	15	2	US-08-661- Sequence 9, Applicatio	1.76e-04
28	96	92.3	19	2	US-08-787- Sequence 41, Applicati	1.76e-04
29	96	92.3	24	3	PCT-US92-0 Sequence 31, Applicati	1.76e-04
30	96	92.3	27	3	PCT-US95-0 Sequence 111, Applicat	1.76e-04
31	96	92.3	27	3	PCT-US95-0 Sequence 112, Applicat	1.76e-04
32	96	92.3	27	3	PCT-US92-0 Sequence 27, Applicati	1.76e-04
33	96	92.3	27	3	PCT-US92-0 Sequence 28, Applicati	1.76e-04
34	96	92.3	30	3	PCT-US92-0 Sequence 26, Applicati	1.76e-04
35	96	92.3	30	3	PCT-US92-0 Sequence 29, Applicati	1.76e-04
36	96	92.3	32	3	US-08-186- Sequence 9, Applicatio	1.76e-04
37	90	86.5	13	3	US-08-787- Sequence 42, Applicati	1.02e-03
38	85	81.7	13	3	PCT-US94-1 Sequence 26, Applicati	4.36e-03
39	55	52.9	91	2	US-08-479- Sequence 10, Applicati	1.72e-01
40	52	50.0	91	2	US-08-308- Sequence 5, Applicatio	3.72e+01
41	52	50.0	91	2	US-08-479- Sequence 9, Applicatio	3.72e+01
42	52	50.0	92	1	US-08-208- Sequence 20, Applicati	3.72e+01
43	52	50.0	92	1	US-08-539- Sequence 20, Applicati	3.72e+01
44	52	50.0	92	1	US-08-167- Sequence 20, Applicati	3.72e+01
45	51	49.0	388	2	US-08-705- Sequence 4, Applicatio	4.80e+01

ALIGNMENTS

RESULT 1
ID PCT-US93-11703-69 STANDARD; PRI: 15 AA.
XX
AC
XX
XX
DT
XX
DE
XX
XX
Sequence 69, Application PC/TUS9311703
GENERAL INFORMATION:
APPLICANT: Chiron Mimotopes Pty. Ltd.
TITLE OF INVENTION: T-Cell Epitopes
NUMBER OF SEQUENCES: 75
CORRESPONDENCE ADDRESS:
ADDRESSEE: Grant D. Green
STREET: 4560 Horton St.
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/11703
FILING DATE: 28-DEC-1993
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/984,852
FILING DATE: 02-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Green, Grant D.
REGISTRATION NUMBER: 31,259
REFERENCE/DOCKET NUMBER: 0222.101
TELECOMMUNICATION INFORMATION:
TELEPHONE: 510-655-3542
INFORMATION FOR SEQ ID NO: 69:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide

CC TITLE OF INVENTION: Protective 17 kDa Malaria Hepatic and
 CC TITLE OF INVENTION: Erythrocytic stage Immunogen and Gene
 CC NUMBER OF SEQUENCES: 11
 CC CORRESPONDENCE ADDRESS:
 CC ADDRESSEE: Naval Medical R & D Command
 CC STREET: Bldg 1, T-12, 8901 Wisconsin Avenue
 CC CITY: Bethesda
 CC STATE: Maryland
 CC COUNTRY: U.S.A
 CC ZIP: 20889-5606
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Floppy disk
 CC COMPUTER: IBM PC compatible
 CC OPERATING SYSTEM: PC-DOS/MS-DOS
 CC SOFTWARE: Patent in Release #1.0, Version #1.25
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: US/08/319,704
 CC FILING DATE: 07-OCT-1994
 CC CLASSIFICATION: 424
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: A. David Spevack
 CC REGISTRATION NUMBER: 24,743
 CC REFERENCE/DOCKET NUMBER: 75,206
 CC TELECOMMUNICATION INFORMATION:
 CC TELEPHONE: (301) 295-6759
 CC TELEFAX: (301) 295-1022
 CC INFORMATION FOR SEQ ID NO: 10:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 15 amino acids
 CC TYPE: amino acid
 CC TOPOLOGY: linear
 CC MOLECULE TYPE: peptide
 CC SEQUENCE 15 AA; 1725 MW; 1271 CN;
 CC
 CC Query Match 100.0%; Score 104; DB 2; Length 15;
 CC Best Local Similarity 100.0%; Pred. No. 1.64e-05;
 CC Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 Db 1 QYIKANSKEIGITEL 15
 QY 1 QYIKANSKEIGITEL 15
 CC
 CC RESULT 5
 CC ID PCT-US95-13841-7 STANDARD; PRT: 17 AA.
 CC XX xxxxxx
 CC
 CC Sequence 7, Application PC/TUS9513841
 CC
 CC GENERAL INFORMATION:
 CC APPLICANT: United Biomedical Inc; Walfield, Alan M.;
 CC APPLICANT: Wang, Chang Y1
 CC TITLE OF INVENTION: Synthetic Ige Membrane Anchor
 CC TITLE OF INVENTION: Peptide Immunogens for the Treatment of Allergy
 CC NUMBER OF SEQUENCES: 32
 CC CORRESPONDENCE ADDRESS:
 CC ADDRESSEE: Maria C.H. Lin
 CC STREET: 345 Park Avenue
 CC CITY: New York
 CC STATE: NY
 CC COUNTRY: USA
 CC ZIP: 10154
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Floppy disk
 CC COMPUTER: IBM PC compatible
 CC OPERATING SYSTEM: PC-DOS/MS-DOS
 CC SOFTWARE: WordPerfect 5.1
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: PCT/US95/13841

CC FILING DATE: 25-OCT-1995
 CC CLASSIFICATION:
 CC PRIOR APPLICATION DATA:
 CC APPLICATION NUMBER: 08/328,519
 CC FILING DATE: 25-OCT-1994
 CC CLASSIFICATION:
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: Lin, Maria C.H.
 CC REGISTRATION NUMBER: 29,323
 CC REFERENCE/DOCKET NUMBER: 1151-4117
 CC TELECOMMUNICATION INFORMATION:
 CC TELEPHONE: 212-758-4800
 CC TELEFAX: 212-751-6849
 CC TELEX: 421792
 CC INFORMATION FOR SEQ ID NO: 7:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 17 amino acids
 CC TYPE: amino acid
 CC TOPOLOGY: linear
 CC MOLECULE TYPE: peptide
 CC SEQUENCE 17 AA; 1981 MW; 1619 CN;
 CC
 CC Query Match 100.0%; Score 104; DB 3; Length 17;
 CC Best Local Similarity 100.0%; Pred. No. 1.64e-05;
 CC Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 Db 3 QYIKANSKEIGITEL 17
 QY 1 QYIKANSKEIGITEL 15
 CC
 CC RESULT 6
 CC ID PCT-US95-08596-23 STANDARD; PRT: 17 AA.
 CC XX xxxxxx
 CC
 CC Sequence 23, Application PC/TUS9508596
 CC
 CC GENERAL INFORMATION:
 CC APPLICANT:
 CC TITLE OF INVENTION: Proinsulin Peptide Compounds for Detecting
 CC TITLE OF INVENTION: and treating Type I Diabetes
 CC NUMBER OF SEQUENCES: 23
 CC CORRESPONDENCE ADDRESS:
 CC ADDRESSEE: LAHIVE & COCKFIELD
 CC STREET: 60 State Street, suite 510
 CC CITY: Boston
 CC STATE: Massachusetts
 CC COUNTRY: USA
 CC ZIP: 02109-1875
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Floppy disk
 CC COMPUTER: IBM PC compatible
 CC OPERATING SYSTEM: PC-DOS/MS-DOS
 CC SOFTWARE: ASCII Text
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: PCT/US95/08596
 CC FILING DATE:
 CC CLASSIFICATION:
 CC PRIOR APPLICATION DATA:
 CC APPLICATION NUMBER: US 08/272,220
 CC FILING DATE: 08-JULY-1994
 CC CLASSIFICATION:
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: DeConti, Giulio A., Jr.
 CC REGISTRATION NUMBER: 31,503
 CC REFERENCE/DOCKET NUMBER: DCI-092PC
 CC TELECOMMUNICATION INFORMATION:
 CC TELEPHONE: (617)227-7400
 CC TELEFAX: (617)227-5941

CC INFORMATION FOR SEQ ID NO: 23:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 17 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC SEQUENCE 17 AA: 1969 MW: 1620 CN;

Query Match 100.0%; Score 104; DB 3; Length 17;
Best Local Similarity 100.0%; Pred. No. 1.64e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3 QYIKANSKFIGITEL 17
QY 1 QYIKANSKFIGITEL 15

RESULT 7
ID US-08-488-351A-4 STANDARD; PRT; 17 AA.

XX AC xxxxxx

Sequence 4, Application US/08488351A

Sequence 4, Application US/08488351A

Patent No. 5843446

GENERAL INFORMATION:

APPLICANT: Ladd, Anna

APPLICANT: Wang, Chang Yi

APPLICANT: Zamb, Timothy

TITLE OF INVENTION: Immunogenic LHRH peptide constructs

TITLE OF INVENTION: and synthetic universal immune stimulators for vaccines

NUMBER OF SEQUENCES: 114

CORRESPONDENCE ADDRESS:

ADDRESSEE: Maria C.H. Lin

STREET: 345 Park Avenue

CITY: New York

STATE: NY

COUNTRY: US

ZIP: 10154-0053

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/488,351A

FILING DATE: 7-JUN-1995

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/446,692

FILING DATE: 7-JUN-1995

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/229,275

FILING DATE: 14-APR-1994

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/057,166

FILING DATE: 27-APR-1992

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Maria C.H. Lin

REGISTRATION NUMBER: 29,323

REFERENCE/DOCKET NUMBER: 1151-4146 US2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212)415-8745

TELEFAX: (516)751-6849

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

CC LENGTH: 17 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 17 AA: 1981 MW: 1619 CN;

Query Match 100.0%; Score 104; DB 2; Length 17;
Best Local Similarity 100.0%; Pred. No. 1.64e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3 QYIKANSKFIGITEL 17
QY 1 QYIKANSKFIGITEL 15

RESULT 8
ID US-08-446-692-4 STANDARD; PRT; 17 AA.

XX AC xxxxxx

Sequence 4, Application US/08446692

Sequence 4, Application US/08446692

Patent No. 5759551

GENERAL INFORMATION:

APPLICANT: Ladd, Anna

APPLICANT: Wang, Chang Yi

APPLICANT: Zamb, Timothy

TITLE OF INVENTION: Immunogenic LHRH peptide constructs

TITLE OF INVENTION: and synthetic universal immune stimulators for vaccines

NUMBER OF SEQUENCES: 114

CORRESPONDENCE ADDRESS:

ADDRESSEE: Maria C.H. Lin

STREET: 345 Park Avenue

CITY: New York

STATE: NY

COUNTRY: US

ZIP: 10154-0053

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/446,692

FILING DATE: 7-JUN-1995

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Maria C.H. Lin

REGISTRATION NUMBER: 29,323

REFERENCE/DOCKET NUMBER: 1151-4146 US2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212)415-8745

TELEFAX: (516)751-6849

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 17 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

SEQUENCE 17 AA: 1981 MW: 1619 CN;

Query Match 100.0%; Score 104; DB 2; Length 17;
Best Local Similarity 100.0%; Pred. No. 1.64e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3 QYIKANSKFIGITEL 17
QY 1 QYIKANSKFIGITEL 15

```

RESULT 9
ID US-08-445-692-13 STANDARD; PRT; 27 AA.
XX
AC xxxxxx
XX
DI
XX
DE
DE
XX
CC Sequence 13, Application US/08445692
CC
CC Sequence 13, Application US/08445692
CC Patent No. 5759551
CC GENERAL INFORMATION:
CC APPLICANT: Ladd, Anna
CC APPLICANT: Wang, Chang yi
CC APPLICANT: Zamb, Timothy
CC
CC TITLE OF INVENTION: Immunogenic LHRH peptide constructs
CC and synthetic universal immune stimulators for vaccines
CC NUMBER OF SEQUENCES: 114
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Maria C.H. Lin
CC STREET: 345 Park Avenue
CC CITY: New York
CC STATE: NY
CC COUNTRY: US
CC ZIP: 10154-0053
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC FILING DATE: 7-JUN-1995
CC APPLICATION NUMBER: US/08/446,692
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC FILING DATE: 7-JUN-1995
CC APPLICATION NUMBER: 29,323
CC CLASSIFICATION: 424
CC REFERENCE/DOCKET NUMBER: 1151-4146 US2
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212)415-8745
CC TELEFAX: (516)751-6849
CC INFORMATION FOR SEQ ID NO: 13:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 27 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 27 AA; 3165 MW; 4134 CN;
Query Match 100.0%; Score 104; DB 2; Length 27;
Best Local Similarity 100.0%; Pred. No. 1.64e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 3 QYIKANSKFIGHTEL 17
QY 1 QYIKANSKFIGHTEL 15
RESULT 10
ID US-08-488-351A-13 STANDARD; PRT; 27 AA.
XX
AC xxxxxx
XX
DI
XX
DE
DE
XX
CC Sequence 13, Application US/08488351A
CC
CC Sequence 13, Application US/08488351A
CC Patent No. 5843446
CC GENERAL INFORMATION:
CC APPLICANT: Ladd, Anna
CC APPLICANT: Wang, Chang yi
CC APPLICANT: Zamb, Timothy

```

```

CC
CC TITLE OF INVENTION: Immunogenic LHRH peptide constructs
CC and synthetic universal immune stimulators for vaccines
CC NUMBER OF SEQUENCES: 114
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Maria C.H. Lin
CC STREET: 345 Park Avenue
CC CITY: New York
CC STATE: NY
CC COUNTRY: US
CC ZIP: 10154-0053
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC FILING DATE: 7-JUN-1995
CC APPLICATION NUMBER: US/08/488,351A
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC FILING DATE: 7-JUN-1995
CC APPLICATION NUMBER: US 08/446,692
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC FILING DATE: 14-APR-1994
CC APPLICATION NUMBER: US 08/229,275
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC FILING DATE: 27-APR-1992
CC APPLICATION NUMBER: US 08/057,166
CC CLASSIFICATION: 424
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Maria C.H. Lin
CC REGISTRATION NUMBER: 29,323
CC REFERENCE/DOCKET NUMBER: 1151-4146 US2
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212)415-8745
CC TELEFAX: (516)751-6849
CC INFORMATION FOR SEQ ID NO: 13:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 27 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 27 AA; 3165 MW; 4134 CN;
Query Match 100.0%; Score 104; DB 2; Length 27;
Best Local Similarity 100.0%; Pred. No. 1.64e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 3 QYIKANSKFIGHTEL 17
QY 1 QYIKANSKFIGHTEL 15
RESULT 11
ID PCT-US93-11703-63 STANDARD; PRT; 31 AA.
XX
AC xxxxxx
XX
DI
XX
DE
DE
XX
CC Sequence 63, Application PC/TUS9311703
CC
CC Sequence 63, Application PC/TUS9311703
CC GENERAL INFORMATION:
CC APPLICANT: Chiron Mimotopes Pty. Ltd.
CC TITLE OF INVENTION: T-Cell Epitopes
CC NUMBER OF SEQUENCES: 75
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Grant D. Green
CC STREET: 4560 Horton St.
CC CITY: Emeryville

```

CC STATE: CA
CC COUNTRY: USA
CC ZIP: 94608
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.30B
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US93/11703
CC FILING DATE: 28-DEC-1993
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/984,852
CC FILING DATE: 02-DEC-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Green, Grant D.
CC REGISTRATION NUMBER: 31,259
CC REFERENCE/DOCKET NUMBER: 0222.101
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 510-601-2706
CC TELEFAX: 510-655-3542
CC INFORMATION FOR SEQ ID NO: 63:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 31 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 31 AA; 3583 MW; 5387 CN;

Query Match 100.0%; Score 104; DB 3; Length 31;
Best Local Similarity 100.0%; Pred. No. 1.64e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 10 QYIKANSKFIGITEL 24
QY 1 QYIKANSKFIGITEL 15

RESULT 12
ID US-08-446-692-63 STANDARD; PRT; 37 AA.

XX
AC
AC
DT
DT
XX

Sequence 63, Application US/08446692

Sequence 63, Application US/08446692

Patent No. 5759551

GENERAL INFORMATION:

APPLICANT: Ladd, Anna

APPLICANT: Wang, Chang Yi

APPLICANT: Zamb, Timothy

TITLE OF INVENTION: Immunogenic LHRH peptide constructs

TITLE OF INVENTION: and synthetic universal immune stimulators for vaccines

NUMBER OF SEQUENCES: 114

CORRESPONDENCE ADDRESS:

ADDRESSEE: Maria C.H. Lin

STREET: 345 Park Avenue

CITY: New York

STATE: NY

COUNTRY: US

ZIP: 10154-0053

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/446,692

FILING DATE: 7-JUN-1995

CC CLASSIFICATION: 424
CC ATTORNEY/AGENT INFORMATION:

CC NAME: Maria C.H. Lin

CC REGISTRATION NUMBER: 29,323

CC REFERENCE/DOCKET NUMBER: 1151-4146 US2

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (212)415-8745

CC TELEFAX: (516)751-6849

CC INFORMATION FOR SEQ ID NO: 63:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 37 amino acids

CC TYPE: amino acid

CC TOPOLOGY: linear

CC MOLECULE TYPE: peptide

CC SEQUENCE 37 AA; 4060 MW; 8502 CN;

Query Match 100.0%; Score 104; DB 2; Length 37;

Best Local Similarity 100.0%; Pred. No. 1.64e-05;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 5 QYIKANSKFIGITEL 19

QY 1 QYIKANSKFIGITEL 15

RESULT 13

ID US-08-488-351A-57 STANDARD; PRT; 37 AA.

XX
AC
AC
XX

Sequence 57, Application US/08488351A

Sequence 57, Application US/08488351A

Patent No. 5843446

GENERAL INFORMATION:

APPLICANT: Ladd, Anna

APPLICANT: Wang, Chang Yi

APPLICANT: Zamb, Timothy

TITLE OF INVENTION: Immunogenic LHRH peptide constructs

TITLE OF INVENTION: and synthetic universal immune stimulators for vaccines

NUMBER OF SEQUENCES: 114

CORRESPONDENCE ADDRESS:

ADDRESSEE: Maria C.H. Lin

STREET: 345 Park Avenue

CITY: New York

STATE: NY

COUNTRY: US

ZIP: 10154-0053

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/488,351A

FILING DATE: 7-JUN-1995

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/446,692

FILING DATE: 7-JUN-1995

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/229,275

FILING DATE: 14-APR-1994

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/057,166

FILING DATE: 27-APR-1992

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Maria C.H. Lin

CC REGISTRATION NUMBER: 29,323
CC REFERENCE/DOCKET NUMBER: 1151-4146 US2
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212)415-8745
CC TELEFAX: (516)751-6849
CC INFORMATION FOR SEQ ID NO: 57:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 37 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
SQ SEQUENCE 37 AA; 4060 MW; 7526 CN;

Query Match 100.0%; Score 104; DB 2; Length 37;
Best Local Similarity 100.0%; Pred. No. 1.64e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 21 QYIKANSKFIGITEL 35
QY 1 QYIKANSKFIGITEL 15
|||||

RESULT 14

ID US-08-446-692-57 STANDARD; PRT; 37 AA.
XX
AC xxxxxx
XX

Sequence 57, Application US/08446692

Sequence 57, Application US/08446692
Patent No. 5739551

GENERAL INFORMATION:

APPLICANT: Ladd, Anna

APPLICANT: Wang, Chang Yi

APPLICANT: Zamb, Timothy

TITLE OF INVENTION: Immunogenic LHRH peptide constructs

TITLE OF INVENTION: and synthetic universal immune stimulators for vaccines

NUMBER OF SEQUENCES: 114

CORRESPONDENCE ADDRESS:

ADDRESSEE: Maria C.H. Lin

STREET: 345 Park Avenue

CITY: New York

STATE: NY

COUNTRY: US

ZIP: 10154-0053

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/446,692

FILING DATE: 7-JUN-1995

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Maria C.H. Lin

REGISTRATION NUMBER: 29,323

REFERENCE/DOCKET NUMBER: 1151-4146 US2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212)415-8745

TELEFAX: (516)751-6849

INFORMATION FOR SEQ ID NO: 57:

SEQUENCE CHARACTERISTICS:

LENGTH: 37 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

SQ SEQUENCE 37 AA; 4060 MW; 7526 CN;

Query Match 100.0%; Score 104; DB 2; Length 37;
Best Local Similarity 100.0%; Pred. No. 1.64e-05;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 21 QYIKANSKFIGITEL 35
QY 1 QYIKANSKFIGITEL 15
|||||

RESULT 15

ID US-08-488-351A-63 STANDARD; PRT; 37 AA.
XX
AC xxxxxx
XX

Sequence 63, Application US/08488351A

Sequence 63, Application US/08488351A

Patent No. 5843446

GENERAL INFORMATION:

APPLICANT: Ladd, Anna

APPLICANT: Wang, Chang Yi

APPLICANT: Zamb, Timothy

TITLE OF INVENTION: Immunogenic LHRH peptide constructs

TITLE OF INVENTION: and synthetic universal immune stimulators for vaccines

NUMBER OF SEQUENCES: 114

CORRESPONDENCE ADDRESS:

ADDRESSEE: Maria C.H. Lin

STREET: 345 Park Avenue

CITY: New York

STATE: NY

COUNTRY: US

ZIP: 10154-0053

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/488,351A

FILING DATE: 7-JUN-1995

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/446,692

FILING DATE: 7-JUN-1995

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/229,275

FILING DATE: 14-APR-1994

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/057,166

FILING DATE: 27-APR-1992

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Maria C.H. Lin

REGISTRATION NUMBER: 29,323

REFERENCE/DOCKET NUMBER: 1151-4146 US2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212)415-8745

TELEFAX: (516)751-6849

INFORMATION FOR SEQ ID NO: 63:

SEQUENCE CHARACTERISTICS:

LENGTH: 37 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

SQ SEQUENCE 37 AA; 4060 MW; 8502 CN;

Query Match 100.0%; Score 104; DB 2; Length 37;
Best Local Similarity 100.0%; Pred. No. 1.64e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 5 QYIKANSKFIGITEL 19

QY 1 QYKANSKFIGITEL 15

RESULT 16
ID US-08-488-351A-35 STANDARD; PRT; 47 AA.

XX xxxxxx

Sequence 35, Application US/08488351A

Sequence 35, Application US/08488351A

Patent No. 5843446

GENERAL INFORMATION:

APPLICANT: Ladd, Anna

APPLICANT: Wang, Chang Yi

APPLICANT: Zamb, Timothy

TITLE OF INVENTION: Immunogenic LHRH peptide constructs

TITLE OF INVENTION: and synthetic universal immune stimulators for vaccines

NUMBER OF SEQUENCES: 114

CORRESPONDENCE ADDRESS:

STREET: 345 Park Avenue

CITY: New York

STATE: NY

COUNTRY: US

ZIP: 10154-0053

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/488,351A

FILING DATE: 7-JUN-1995

CLASSIFICATION: 424

PRIOR APPLICATION NUMBER: US 08/446,692

FILING DATE: 7-JUN-1995

CLASSIFICATION: 424

APPLICATION NUMBER: US 08/229,275

FILING DATE: 14-APR-1994

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/057,166

FILING DATE: 27-APR-1992

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Maria C.H. Lin

REGISTRATION NUMBER: 29,323

REFERENCE/DOCKET NUMBER: 1151-4146 US2

TELEPHONE: (212)415-8745

TELEFAX: (516)751-6849

INFORMATION FOR SEQ ID NO: 35:

SEQUENCE CHARACTERISTICS:

LENGTH: 47 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

SEQUENCE 47 AA; 5243 MW; 12301 CN;

Query Match 100.0%; Score 104; DB 2; Length 47;
Best Local Similarity 100.0%; Pred. No. 1.64e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 21 QYKANSKFIGITEL 35

XXXXXXXXXXXXXXXXXXXX

QY 1 QYKANSKFIGITEL 15

RESULT 17
ID US-08-446-692-35 STANDARD; PRT; 47 AA.

XX xxxxxx

Sequence 35, Application US/08446692

Sequence 35, Application US/08446692

Patent No. 5759551

GENERAL INFORMATION:

APPLICANT: Ladd, Anna

APPLICANT: Wang, Chang Yi

APPLICANT: Zamb, Timothy

TITLE OF INVENTION: Immunogenic LHRH peptide constructs

TITLE OF INVENTION: and synthetic universal immune stimulators for vaccines

NUMBER OF SEQUENCES: 114

CORRESPONDENCE ADDRESS:

STREET: 345 Park Avenue

CITY: New York

STATE: NY

COUNTRY: US

ZIP: 10154-0053

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/446,692

FILING DATE: 7-JUN-1995

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Maria C.H. Lin

REGISTRATION NUMBER: 29,323

REFERENCE/DOCKET NUMBER: 1151-4146 US2

TELEPHONE: (212)415-8745

TELEFAX: (516)751-6849

INFORMATION FOR SEQ ID NO: 35:

SEQUENCE CHARACTERISTICS:

LENGTH: 47 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

SEQUENCE 47 AA; 5243 MW; 12301 CN;

Query Match 100.0%; Score 104; DB 2; Length 47;
Best Local Similarity 100.0%; Pred. No. 1.64e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 21 QYKANSKFIGITEL 35

XXXXXXXXXXXXXXXXXXXX

QY 1 QYKANSKFIGITEL 15

Search completed: Tue Aug 17 16:07:13 1999
Job time : 9 secs.

 W O R L D

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Aug 17 15:53:45 1999; MasPar time 4.57 Seconds

Tabular output not generated. 184.098 Million cell updates/sec

Title: >US-09-049-847-2
 Description: (1-21) from US09049847.ppe
 Perfect Score: 158
 Sequence: 1 FNNFTVFWLRVPKVSASHLE 21

Scoring table: PAM 150
 Gap 15

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0*
 Listing first 45 summaries

Database: pir60

Statistics: 1:pir1 2:pir2 3:pir3 4:pir4

Mean 29.518; Variance 48.854; scale 0.604

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	158	100.0	1315	1 BTCLTN	tentoxylisin (EC 3.4.	4.98e-18
2	95	60.1	1268	2 S33411	botulinum neurotoxin	3.74e-05
3	95	60.1	1297	2 S39791	neurotoxin - Clostrid	3.74e-05
4	93	58.9	366	2 S48110	neurotoxin type F - C	8.82e-05
5	93	58.9	369	2 S48109	neurotoxin type F - C	8.82e-05
6	93	58.9	1274	2 I40813	neurotoxin type F - C	8.82e-05
7	93	58.9	1291	2 I40631	non-proteolytic botul	8.82e-05
8	93	58.9	1291	2 I40840	botulinum neurotoxin	8.82e-05
9	90	57.0	1296	1 BTCLAB	botulinum neurotoxin	8.82e-05
10	88	55.7	357	2 S48106	neurotoxin type E - C	7.30e-04
11	88	55.7	1251	2 JH0257	botulinum neurotoxin	7.30e-04
12	88	55.7	1251	2 JH0256	botulinum neurotoxin	7.30e-04
13	88	55.7	1252	2 S21178	botulinum neurotoxin	7.30e-04
14	85	53.8	1296	2 I40645	botulinum neurotoxin	2.53e-03
15	84	53.2	464	1 MNVUWC	nonstructural protein	3.82e-03
16	84	53.2	467	1 MNVUW1	nonstructural protein	3.82e-03
17	81	51.3	1285	2 S70582	botulinum neurotoxin	1.29e-02
18	81	51.3	1291	2 A49777	botulinum neurotoxin	1.29e-02
19	81	51.3	1291	2 S46431	botulinum neurotoxin	1.29e-02
20	77	48.7	449	2 S23158	nucleocapsid protein	6.32e-02
21	76	48.1	1196	2 S46430	botulinum neurotoxin-	9.34e-02
22	76	48.1	1196	2 JQ1467	toxin, nontoxic compo	9.34e-02
23	72	45.6	406	2 F70930	hypothetical protein	4.34e-01

ALIGNMENTS

RESULT ENTRY	1	BTCLIN	#type complete
TITLE	1162	tentoxylisin (EC 3.4.24.68)	precursor - Clostridium tetani
ALTERNATE_NAMES	1162	tetanus neurotoxin	
ORGANISM	1162	#formal name Clostridium tetani	
DATE	1162	31-Mar-1988	#sequence revision 31-Mar-1988 #text_change
ACCESSIONS	1162	A25689; A25757; A25194; B25194; A60759; S69348; S09364	
REFERENCE	1162	A25689	
#authors	1162	Eisel, U.; Jarausch, W.; Goretzki, K.; Henschel, A.; Engels, J.; Weller, U.; Hude, M.; Habermann, E.; Niemann, H.	
#journal	1162	EMBO J. (1986) 5:2495-2502	
#title	1162	Tetanus toxin: primary structure, expression in E. coli, and homology with botulinum toxins.	
#cross-references	1162	MUID:87053814	
#accession	1162	A25689	
#molecule_type	1162	DNA	
#residues	1162	1-1315	#label EIS
#cross-references	1162	GB:X04436; NID:g40769; PID:g40770	
REFERENCE	1162	A25757	
#authors	1162	Fairweather, N.F.; Lyness, V.A.	
#journal	1162	Nucleic Acids Res. (1986) 14:7809-7812	
#title	1162	The complete nucleotide sequence of tetanus toxin.	
#cross-references	1162	MUID:87040747	
#accession	1162	A25757	
#molecule_type	1162	DNA	
#residues	1162	1-1315	#label FAI
#cross-references	1162	GB:X06214; NID:g40773; PID:g40774	
#experimental_source	1162	strain CN3911	
REFERENCE	1162	A25194	
#authors	1162	Fairweather, N.F.; Lyness, V.A.; Pickard, D.J.; Allen, G.; Thomson, R.O.	
#journal	1162	J. Bacteriol. (1986) 165:21-27	
#title	1162	Cloning, nucleotide sequencing, and expression of tetanus toxin fragment C in Escherichia coli.	
#cross-references	1162	MUID:86085672	
#accession	1162	A25194	
#molecule_type	1162	DNA	
#residues	1162	743-1315	#label FA2
#cross-references	1162	GB:M12739; NID:g144920; PID:g144921	
#accession	1162	B25194	
#molecule_type	1162	protein	
#residues	1162	865-894	#label FA3
REFERENCE	1162	A60759	
#authors	1162	Matsuda, M.; Lei, D.L.; Sugimoto, N.; Ozutsumi, K.; Okabe, T.	
#journal	1162	Infect. Immun. (1989) 57:3588-3593	
#title	1162	Isolation, purification, and characterization of fragment B,	

24	72	45.6	1276	2	S11455	botulinum neurotoxin	4.34e-01
25	70	44.3	1162	2	T40817	botulinum toxin nont	9.20e-01
26	70	44.3	1162	2	A47708	progenitor toxin nont	9.20e-01
27	68	43.0	699	2	I38073	nucleolar phosphoprot	1.92e-00
28	67	42.4	152	2	D64943	probable membrane pro	2.77e-00
29	67	42.4	789	1	QXBY32	gene cox1 intron 2 pr	2.77e-00
30	67	42.4	854	2	S78644	gene cox1 intron 2 pr	2.77e-00
31	66	41.8	241	2	S17822	spheroidene monooxyge	3.97e-00
32	66	41.8	244	2	H70591	hypothetical protein	3.97e-00
33	66	41.8	591	2	S04401	spheroidene monooxyge	3.97e-00
34	66	41.8	1165	2	I40644	botulinum neurotoxin	3.97e-00
35	66	41.8	1193	2	S68218	botulinum neurotoxin	3.97e-00
36	66	41.8	1193	2	JC4901	botulinum neurotoxin	3.97e-00
37	65	41.1	203	2	A36886	surface protein PAg n	5.67e-00
38	65	41.1	836	2	F71546	probable DNA gyrase c	5.67e-00
39	65	41.1	854	2	S17995	gene COX1 intron 1 pr	5.67e-00
40	63	39.9	278	2	T00613	hypothetical protein	1.15e-01
41	63	39.9	547	2	S51475	cytochrome p450 - mai	1.15e-01
42	63	39.9	605	1	ABPGS	serum albumin precurs	1.15e-01
43	63	39.9	728	2	S57142	hypothetical protein	1.15e-01
44	63	39.9	3972	2	S75251	hypothetical protein	1.15e-01
45	62	39.2	380	2	S33720	NS1 protein - dengue	1.62e-01


```

ACCESSIONS: * S48110
REFERENCE
#authors Campbell, K.D.; Collins, M.D.; East, A.K.
#journal J. Clin. Microbiol. (1993) 31:2255-2262
#title Gene probes for identification of the botulin neurotoxin
#note gene and specific identification of neurotoxin types B, E,
and F.
#cross-references MUID:94013372
#accession S48110
##status preliminary; translation not shown
##molecule_type DNA
##residues 1-366 #label CAM
##cross-references EMBL:X70821; NID:g407792; PID:g407793
CLASSIFICATION #superfamily tetanus toxin
KEYWORDS neurotoxin
SUMMARY #length 366 #checksum 556

Query Match 58.9%; Score 93; DB 2; Length 366;
Best Local Similarity 57.1%; Pred. No. 8.82e-05;
Matches 8; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 297 YONFISFWVRPK 310
:|||||:|
QY 1 FNNFTVSFWLRVPK 14

RESULT 5
ENTRY #type fragment
TITLE neurotoxin type F - Clostridium botulinum (fragment)
ORGANISM #formal name Clostridium botulinum
DATE 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change
SUMMARY #length 369 #checksum 5830

ACCESSIONS
REFERENCE S48109
#authors Campbell, K.D.; Collins, M.D.; East, A.K.
#journal J. Clin. Microbiol. (1993) 31:2255-2262
#title Gene probes for identification of the botulin neurotoxin
#note gene and specific identification of neurotoxin types B, E,
and F.
#cross-references MUID:94013372
#accession S48109
##status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-369 #label CAM
##cross-references EMBL:X70820; NID:g407790; PID:g407791
##note the nucleotide sequence was submitted to the EMBL Data
Library, January 1993
CLASSIFICATION #superfamily tetanus toxin
SUMMARY #length 369 #checksum 5830

Query Match 58.9%; Score 93; DB 2; Length 369;
Best Local Similarity 57.1%; Pred. No. 8.82e-05;
Matches 8; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 297 YONFISFWVRPK 310
:|||||:|
QY 1 FNNFTVSFWLRVPK 14

RESULT 6
ENTRY #type complete
TITLE neurotoxin type F - Clostridium botulinum
ORGANISM #formal name Clostridium botulinum
DATE 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change
SUMMARY #length 369 #checksum 5830

ACCESSIONS
REFERENCE S48108
#authors East, A.K.; Richardson, P.T.; Allaway, D.; Collins, M.D.;
Roberts, I.A.; Thompson, D.E.
#journal FEMS Microbiol. Lett. (1992) 96:225-230
#title Sequence of the gene encoding type F neurotoxin of
Clostridium botulinum.

```

```

#accession I40813
##status preliminary; translated from GB/EMBL/DBJ
##molecule_type DNA
##residues 1-1274 #label RES
##cross-references GB:M92906; NID:g144866; PID:g144867
REFERENCE S48103
#authors Campbell, K.D.; Collins, M.D.; East, A.K.
#journal J. Clin. Microbiol. (1993) 31:2255-2262
#title Gene probes for identification of the botulin neurotoxin
#note gene and specific identification of neurotoxin types B, E,
and F.
#cross-references MUID:94013372
#accession S48108
##status preliminary; translation not shown
##molecule_type DNA
##residues 634-1002 #label CAM
##cross-references EMBL:X70816; NID:g407788; PID:g407789
CLASSIFICATION #superfamily tetanus toxin
KEYWORDS neurotoxin
SUMMARY #length 1274 #molecular-weight 146708 #checksum 2696

Query Match 58.9%; Score 93; DB 2; Length 1274;
Best Local Similarity 57.1%; Pred. No. 8.82e-05;
Matches 8; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 930 YONFISFWVRPK 943
:|||||:|
QY 1 FNNFTVSFWLRVPK 14

RESULT 7
ENTRY #type complete
TITLE non-proteolytic botulinum neurotoxin type B precursor -
Clostridium botulinum
ORGANISM #formal name Clostridium botulinum
DATE 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change
SUMMARY #length 1274 #molecular-weight 146708 #checksum 2696

ACCESSIONS
REFERENCE I40631
#authors Hutson, R.A.; Collins, M.D.; East, A.K.; Thompson, D.E.
#journal Curr. Microbiol. (1994) 28:101-110
#title Nucleotide sequence of the gene coding for non-proteolytic
Clostridium botulinum type B neurotoxin: comparison with
other clostridial neurotoxins.
#cross-references MUID:94122659
#accession I40631
##status preliminary; translated from GB/EMBL/DBJ
##molecule_type DNA
##residues 1-1291 #label RES
##cross-references EMBL:X71343; NID:g296148; PID:g296149
REFERENCE S48103
#authors Campbell, K.D.; Collins, M.D.; East, A.K.
#journal J. Clin. Microbiol. (1993) 31:2255-2262
#title Gene probes for identification of the botulin neurotoxin
#note gene and specific identification of neurotoxin types B, E,
and F.
#cross-references MUID:94013372
#accession S48103
##status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 634-761, 'E', 763-841, 'M', 843, 'T', 845, 'N', 847-994 #label
CAM1
##cross-references EMBL:X70814; NID:g407778; PID:g407779
##experimental_source non-proteolytic strain 21298 (Scott)
##note the nucleotide sequence was submitted to the EMBL Data
Library, January 1993
#accession S48104
##status preliminary
##molecule_type DNA
##residues 634-843, 'T', 845, 'N', 847-994 #label CAM2
##cross-references EMBL:X70819; NID:g407780; PID:g407781
##experimental_source non-proteolytic strain Eklund 2B (Colworth 229)

```

```

COMMENT Botulinum neurotoxin type B in these strains may possess a capable
catalytic site but are nonetheless inactive.

GENETICS
#gene bont/b
CLASSIFICATION #superfamily tetanus toxin
KEYWORDS metalloprotein; neurotoxin; transmembrane protein; zinc
FEATURE
2-441 #product botulinum neurotoxin type B light chain #status
predicted #label LGH\
#product botulinum neurotoxin type B heavy chain #status
predicted #label HV\
230,234 #binding_site zinc (His) #status predicted\
231 #active_site Glu #status predicted
SUMMARY #length 1291 #molecular-weight 150513 #checksum 5240

Query Match 58.9%; Score 93; DB 2; Length 1291;
Best Local Similarity 64.3%; Pred. No. 8,82e-05;
Matches 9; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 923 ELDFSVFWIRPK 936
QY 1 FNNFTVSWLRVFK 14
|:|:|||||:|:|

RESULT 8
ENTRY A48940 #type complete
TITLE Bontoxilysin (EC 3.4.24.69) B precursor - Clostridium
botulinum
ALTERNATE_NAMES botulinum neurotoxin type B (BONT/B)
ORGANISM Clostridium botulinum
DATE 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change
15-May-1998
ACCESSIONS A48940; S48105; S21575; A42871; S07155; S08562; S07128;
S08573; S08574
REFERENCE A48940
#authors Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Brehm, J.K.;
Atkinson, T.; Minton, N.P.
#journal Appl. Environ. Microbiol. (1992) 58:2345-2354
#title Molecular cloning of the clostridium botulinum structural
gene encoding the type B neurotoxin and determination of
its entire nucleotide sequence.
#cross-references MUID:92384550
#accession A48940
#status preliminary
#molecule_type DNA
#residues 1-1291 #label WHE
#cross-references GB:M01186; NID:g144734; PID:g144735
#experimental_source type B, Danish
#note sequence extracted from NCBI backbone (NCBIN:112080,
NCBIP:112081); this publication is not cited in
GenBank entry CLOBOTB, release 103

REFERENCE S48103
#authors Campbell, K.D.; Collins, M.D.; East, A.K.
#journal J. Clin. Microbiol. (1993) 31:2255-2262
#title Gene probes for identification of the botulin neurotoxin
gene and specific identification of neurotoxin types B, E,
and F.
#cross-references MUID:94013372
#accession S48105
#status preliminary
#molecule_type DNA
#residues 634-994 #label CAM
#cross-references EMBL:X70817; NID:g407782; PID:g407783
#experimental_source proteolytic type B, strain NCTC 7273
REFERENCE S21575
#authors Szabo, E.A.; Pemberton, J.M.; Desmarchelier, P.M.
#submission Submitted to the EMBL Data Library, April 1992
#description Partial amino acid sequence of botulinum neurotoxin type B
and comparison to other Clostridial neurotoxins.
#accession S21575
#molecule_type DNA
#residues 36-217, 'G', 219-224, 'S', 226-246 #label SZA
#cross-references EMBL:X11934; NID:g40383; PID:g40384

REFERENCE A42871
#authors Kurazono, H.; Mochida, S.; Binz, T.; Eisel, U.; Quanz, M.;
Greibenstein, O.; Wernars, K.; Poulain, B.; Tauc, L.;
Niemann, H.
#journal J. Biol. Chem. (1992) 267:14721-14729
#title Minimal essential domains specifying toxicity of the light
chains of tetanus toxin and botulinum neurotoxin type A.
#cross-references MUID:92340509
#accession A42871
#status nucleic acid sequence not shown
#molecule_type mRNA
#residues 1-313, 'S', 315-451 #label KUR
#experimental_source strain Okra
#note sequence extracted from NCBI backbone (NCBIP:109365)
REFERENCE S07155
#authors DasGupta, B.R.; Datta, A.
#journal Biochimie (1988) 70:811-817
#title Botulinum neurotoxin type B (strain 657): partial sequence
and similarity with tetanus toxin.
#cross-references MUID:89000987
#accession S07155
#molecule_type protein
#residues 2-29, 'M', 31-45 #label DAS
#accession S08562
#molecule_type protein
#residues 442-463, 'R', 465-467 #label DA2
REFERENCE S07128
#authors Schmidt, J.J.; Sathyanarathy, V.; DasGupta, B.R.
#journal Arch. Biochem. Biophys. (1985) 238:544-548
#title Partial amino acid sequences of botulinum neurotoxins types B
and E.
#cross-references MUID:85197963
#accession S07128
#status preliminary
#molecule_type protein
#residues 2-16 #label SCH1
#accession S08573
#status preliminary
#molecule_type protein
#residues 2-17 #label SCH2
#accession S08574
#status preliminary
#molecule_type protein
#residues 442-459 #label SCH3
REFERENCE S27125
#authors Schiavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de
Laureto, P.P.; DasGupta, B.R.; Montecucco, C.
#journal Nature (1992) 359:832-835
#title Tetanus and botulinum-B neurotoxins block neurotransmitter
release by proteolytic cleavage of synaptobrevin.
#cross-references MUID:93063293
#contents annotation
COMMENT Botulinum neurotoxins inhibit neurotransmitter release from
cholinergic synapses. This toxin is activated by cleavage into
two chains linked by a disulfide bond.

GENETICS
#gene bont/b
FUNCTION
#description catalyzes hydrolysis of a Gln-Phe peptide bond in
synaptobrevin 2
CLASSIFICATION #superfamily tetanus toxin
KEYWORDS hydrolase; metalloproteinase; neurotoxin; transmembrane
protein; zinc
FEATURE
2-441 #product bontoxilysin B light chain #status experimental
#label LGH\
442-1291 #product bontoxilysin B heavy chain #status experimental
#label HV\
230,234 #binding_site zinc (His) #status predicted\
231 #active_site Glu #status predicted
SUMMARY #length 1291 #molecular-weight 150801 #checksum 9744

Query Match 58.9%; Score 93; DB 1; Length 1291;

```

```

Best Local Similarity 64.3%; Pred. No. 8,82e-05;
Matches 9; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 923 PLDFSVSFWRIPK 936
   1 FNNFTVSFWLRVPK 14

RESULT 9
ENTRY 9
TITLE 9
BTCLAB #type complete
botoxylisin (EC 3.4.24.69) A precursor - Clostridium
botulinum
ALTERNATE_NAMES #formal name Clostridium botulinum
ORGANISM #formal name Clostridium botulinum
DATE 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change
15-May-1998
ACCESSIONS A35294; S09492; S68220; A33401; A53884; A60025; A27000
REFERENCE A35294
#authors Binz, T.; Kurazono, H.; Wille, M.; Prevert, J.; Wernars, K.;
Niemann, H.
#journal J. Biol. Chem. (1990) 265:9153-9158
#title The complete sequence of botulinum neurotoxin type A and
comparison with other clostridial neurotoxins.
#cross-references MUID:90264400
#accession A35294
#molecule_type DNA
#residues 1-1296 #label BIN
#cross-references GB:M30196; NID:g144864; PID:g144865
#experimental_source strain 62A, subtype A
REFERENCE S09492
#authors Thompson, D.E.; Brehm, J.K.; Oultram, J.D.; Swinfield, T.J.;
Shone, C.C.; Atkinson, T.; Melling, J.; Minton, N.P.
#journal Eur. J. Biochem. (1990) 189:73-81
#title The complete amino acid sequence of the Clostridium botulinum
type A neurotoxin, deduced by nucleotide sequence analysis
of the encoding gene.
#cross-references MUID:90235864
#accession S09492
#molecule_type DNA
#residues 1, 'Q', 3-26, 'V', 28-1296 #label THO
#cross-references EMBL:X52065; NID:g40381; PID:g40382
#experimental_source NCTC 2916
REFERENCE S67968
#authors Fujita, R.; Fujinaga, Y.; Inoue, K.; Nakajima, H.; Kumon, H.;
Oguma, K.
#journal FEBS Lett. (1995) 376:41-44
#title Molecular characterization of two forms of
nontoxic-nonhemagglutinin components of Clostridium
botulinum type A progenitor toxins.
#cross-references MUID:96095783
#accession S68220
#status preliminary
#molecule_type DNA
#residues 1-12 #label FUJ
#cross-references EMBL:D67030; DBJ:D50421; NID:g2160224
REFERENCE A33401
#authors Betley, M.J.; Somers, E.; DasGupta, B.R.
#journal Biochem. Biophys. Res. Commun. (1989) 162:1388-1395
#title Characterization of botulinum type A neurotoxin gene:
delineation of the N-terminal encoding region.
#cross-references MUID:89350959
#accession A33401
#molecule_type DNA
#residues 1-35 #label BET
#cross-references GB:M27892; NID:g144880; PID:g551776
REFERENCE A53884
#authors Gimenez, J.A.; DasGupta, B.R.
#journal J. Protein Chem. (1993) 12:351-363
#title Botulinum type A neurotoxin digested with pepsin yields 132,
97, 72, 45, 42, and 18 kD fragments.
#cross-references MUID:94000342
#accession A53884
#status preliminary

```

```

#molecule_type protein
#residues 867-880:1148-1217, 'Y', 1219 #label GIM
#experimental_source strain Hall
#note sequence extracted from NCBI backbone (NCBIP:139159);
sequence modified after extraction from NCBI backbone
REFERENCE A60025
#authors DasGupta, B.R.; Dekleva, M.L.
#journal Biochimie (1990) 72:661-664
#title Botulinum neurotoxin type A: sequence of amino acids at the
N-terminus and around the nicking site.
#cross-references MUID:91120847
#accession A60025
#molecule_type protein
#residues 2-6:445-453, 'X', 455-457 #label DAS1
REFERENCE A27000
#authors DasGupta, B.R.; Foley, J.; Niece, R.
#journal Biochemistry (1987) 26:4162
#title Partial sequence of the light chain of botulinum neurotoxin
type A.
#accession A27000
#molecule_type protein
#residues 2-47 #label DAS2
REFERENCE A49708
#authors Binz, T.; Blasi, J.; Yamasaki, S.; Baumeister, A.; Link, E.;
Suedhof, T.C.; Jahn, R.; Niemann, H.
#journal J. Biol. Chem. (1994) 269:1617-1620
#title Proteolysis of SNAP-25 by types E and A botulinum
neurotoxins.
#cross-references MUID:94124495
#contents annotation
COMMENT Botulinum neurotoxins inhibit neurotransmitter release from
cholinergic synapses. This toxin is activated by cleavage into
two chains linked by a disulfide bond.
GENETICS atx; botA
#gene
#FUNCTION
#description catalyzes hydrolysis of an Asn-Arg peptide bond in
synaptosomal-associated 25K protein (SNAP-25)
CLASSIFICATION #superfamily tetanus toxin
KEYWORDS disulfide bond; hydrolase; metalloproteinase; neurotoxin;
transmembrane protein; zinc
FEATURE
2-444
#product botoxylisin A light chain #status experimental
#label LGHV\
445-1296 #product botoxylisin A heavy chain #status experimental
#label HVV\
223,227 #binding_site zinc (His) #status predicted\
224 #active_site Glu #status predicted\
SUMMARY #length 1296 #molecular-weight 149425 #checksum 7102
Query Match 57.0%; Score 90; DB 1; Length 1296;
Best Local Similarity 57.1%; Pred. No. 3.15e-04;
Matches 8; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
Db 938 YENFTSFWRIPK 951
   1 FNNFTVSFWLRVPK 14
   :::::|:::|:::|
RESULT 10
ENTRY 10
TITLE 10
ORGANISM
DATE
S48106 #type fragment
neurotoxin type E - Clostridium botulinum (fragment)
#formal_name Clostridium botulinum
14-Jul-1995 #sequence_revision 10-Nov-1995 #text_change
12-Jun-1998
S48106
S48103
#accessions
REFERENCE
#authors Campbell, K.D.; Collins, M.D.; East, A.K.
#journal J. Clin. Microbiol. (1993) 31:2255-2262
#title Gene probes for identification of the botulinum neurotoxin
gene and specific identification of neurotoxin types B, E,
and F.
#cross-references MUID:94013372

```

```

#accession      S48106
##status        preliminary; nucleic acid sequence not shown;
                translation not shown
##molecule_type DNA
##residues      1-367 #label CAM
##cross-references EMBL:X70818; NID:g407784; PID:g407785
##note          the nucleotide sequence was submitted to the EMBL Data
                Library, January 1993
CLASSIFICATION  #superfamily tetanus toxin
KEYWORDS         neurotoxin
SUMMARY          #length 367 #checksum 184

Query Match      55.7%; Score 88; DB 2; Length 367;
Best local Similarity 50.0%; Pred. No. 7.30e-04;
Matches          7; Conservative 7; Mismatches 0; Indels 0; Gaps 0;

Db    297 YKFSISFWVRPN 310
      :|||:|||:|:|:|:
QY    1 FNNFTVSEWLRVPK 14

RESULT 11
ENTRY   JH0257           #type complete
TITLE   botulinum neurotoxin type E precursor - Clostridium botulinum
                (strain Beluga)
ORGANISM
DATE    30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change
                08-Sep-1997
ACCESSIONS JH0257; B35294; A50027
REFERENCE   JH0256
AUTHORS     Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.
JOURNAL     Biochem. Biophys. Res. Commun. (1992) 183:107-113
TITLES      Sequences of the botlinal neurotoxin E derived from
                Clostridium botulinum type E (strain Beluga) and
                Clostridium botryticum (strains ATCC 43181 and ATCC 43755).
CROSS-REFERENCES MUID:92181428
#accession      JH0257           nucleic acid sequence not shown
##molecule_type DNA
##residues      1-1251 #label POU
##cross-references EMBL:X52089; NID:g40393; PID:g40394
REFERENCE       B35294
AUTHORS         Binz, T.; Kurazono, H.; Wille, M.; Prevert, J.; Wernars, K.;
                Niemann, H.
JOURNAL         J. Biol. Chem. (1990) 265:9153-9158
TITLES          The complete sequence of botulinum neurotoxin type A and
                comparison with other clostridial neurotoxins.
CROSS-REFERENCES MUID:90264400
#accession      B35294
##status        not compared with conceptual translation
##molecule_type DNA
##residues      1-197,'S',199-252 #label BIN
REFERENCE       A60027
AUTHORS         Gimenez, J.A.; DasGupta, B.R.
JOURNAL         Biochimie (1990) 72:213-217
TITLES          Botulinum neurotoxin type E fragmented with endoproteinase
                Lys-C reveals the site trypsin nicks and homology with
                tetanus neurotoxin.
CROSS-REFERENCES MUID:90344918
#accession      A60027
##molecule_type protein
##residues      420-427 #label GIM
##note          this fragment was generated by proteolysis with Lys-C
                rather than with trypsin
COMMENT         The clostridial neurotoxins are highly potent protein toxins that
                inhibit neurotransmitter release at various synapses.
COMMENT         The heavy chain mediates the binding of toxin to cell receptors
                while the light chain appears to enter target cells.
CLASSIFICATION  #superfamily tetanus toxin
KEYWORDS         neurotoxin
FEATURE         #product botulinum neurotoxin light chain (toxin
                fragment A); status predicted #label LIG

```

Search completed: Tue Aug 17 15:54:11 1999
Job time : 26 secs.

This Page Blank (uspto)

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Aug 17 15:54:28 1999; MacPar time 3.41 Seconds

Tabular output not generated. 173.887 Million cell updates/sec

Title: >US-09-049-847-2
Description: (1-21) from US09049847.pep
Perfect Score: 158
Sequence: 1 FNNFTVFWLVRPKVSASHLE 21

Scoring table: PAM 150
Gap 15

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
l:swissprot

Statistics: Mean 30.150; Variance 44.662; scale 0.675

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	158	100.0	1314	1	TETX_CLOTE	2.29e-20
2	95	60.1	1296	1	BXG_CLOBO	4.20e-06
3	93	58.9	1274	1	BXF_CLOBO	1.08e-05
4	93	58.9	1290	1	BXB_CLOBO	1.08e-05
5	90	57.0	1295	1	BXA_CLOBO	1.08e-05
6	88	55.7	1250	1	BXE_CLOBO	1.11e-04
7	88	55.7	1250	1	BXE_CLOBO	1.11e-04
8	87	55.1	1051	1	VP2_AHSV6	1.76e-04
9	84	53.2	464	1	VNSS_TSWV1	6.88e-04
10	84	53.2	467	1	VNSS_TSWV1	6.88e-04
11	81	51.3	1290	1	BXC1_CLOBO	2.63e-03
12	77	48.7	449	1	VNSS_INSYN	1.51e-02
13	76	48.1	1196	1	BXCN_CLOBO	2.32e-02
14	72	45.6	1276	1	BXD_CLOBO	1.25e-01
15	70	44.3	1162	1	BXEN_CLOBO	2.84e-01
16	70	44.3	1162	1	BXEN_CLOBO	2.84e-01
17	67	42.4	789	1	A12M_YEAST	9.49e-01
18	66	41.8	241	1	CR1A_RHOCA	1.41e-01
19	63	39.9	208	1	FTSQ_STRGR	4.48e-01
20	63	39.9	547	1	CF78_MALZE	4.48e-01
21	63	39.9	605	1	ALBU_PIG	4.48e-01
22	63	39.9	728	1	Y389_YEAST	4.48e-01
23	52	39.2	581	1	AMY1_SCHPO	6.53e-01

RESULT 1

ID	TETX_CLOTE	STANDARD;	PRT;	1314 AA.
AC	P04958;			
DT	13-AUG-1987 (REL. 05, CREATED)			
DT	13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)			
DT	15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)			
DE	TETANUS TOXIN PRECURSOR (EC 3.4.24.68) (TENTOXILYSIN).			
OS	CLOSTRIDIUM TETANI.			
OG	PLASMID.			
OC	BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;			
OC	CLOSTRIDIUM.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 87053814.			
RA	EISEL U., JARASCH W., GORETZKI K., HENSCHEN A., ENGELS J.,			
RA	WELLER U., HUDEL M., HABERMANN E., NIEMANN H.;			
RT	"Tetanus toxin: primary structure, expression in E. coli, and			
RT	homology with botulinum toxins.";			
RL	EMBO J. 5:2495-2502(1986).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CN3911;			
RX	MEDLINE; 87040747.			
RA	FAIRWEATHER N.F., LYNESS V.A.;			
RT	"The complete nucleotide sequence of tetanus toxin.";			
RL	NUCLEIC ACIDS RES. 14:7809-7812(1986).			
RN	[3]			
RP	SEQUENCE OF 742-1314 FROM N.A.			
RX	MEDLINE; 86085672.			
RA	FAIRWEATHER N.F., LYNESS V.A., PICKARD D.J., ALLEN C., THOMSON R.O.;			
RT	"Cloning, nucleotide sequencing, and expression of tetanus toxin			
RT	fragment C in Escherichia coli.";			
RL	J. BACTERIOL. 165:21-27(1986).			
RN	[4]			
RP	PARTIAL SEQUENCE, AND DISULFIDE BONDS.			
RX	MEDLINE; 90201034.			
RA	KRIEGLSTEIN K., HENSCHEN A., WELLER U., HABERMANN E.;			
RT	"Arrangement of disulfide bridges and positions of sulphhydryl groups			
RT	in tetanus toxin.";			
RL	EUR. J. BIOCHEM. 188:39-45(1990).			
RN	[5]			
RP	PARTIAL SEQUENCE.			
RX	MEDLINE; 92037649.			
RA	KRIEGLSTEIN K.G., HENSCHEN A.H., WELLER U., HABERMANN E.;			
RT	"Limited proteolysis of tetanus toxin. Relation to activity and			
RT	identification of cleavage sites.";			
RL	EUR. J. BIOCHEM. 202:41-51(1991).			

RN IDENTIFICATION AS ZINC-PROTEASE.
 RP MEDLINE; 93010948.
 RX SCHIAVO G., POULAIN B., ROSETTO O., BENFENATI F., TAUC L.,
 RA MONTEUCCO C.;
 RT "Tetanus toxin is a zinc protein and its inhibition of
 RT neurotransmitter release and protease activity depend on zinc.";
 RL EMBO J. 11:3577-3583(1992).
 [17]
 RN IDENTIFICATION OF SUBSTRATE.
 RP MEDLINE; 93063293.
 RX SCHIAVO G., BENFENATI F., POULAIN B., ROSETTO O., DE LAURETO P.P.,
 RA DASGUPTA B.R., MONTEUCCO C.;
 RT "Tetanus and botulinum-B neurotoxins block neurotransmitter release
 RT by proteolytic cleavage of synaptobrevin.";
 RL NATURE 359:832-835(1992).
 [18]
 RN X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 874-1314.
 RP MEDLINE; 97475217.
 RX UMLAND T.C., WINGERT L.M., SWAMINATHAN S., FUREY W.F., SCHMIDT J.J.,
 RA SAX M.;
 RT "Structure of the receptor binding fragment HC of tetanus
 RT neurotoxin.";
 RL NAT. STRUCT. BIOL. 4:788-792(1997).
 CC -!- FUNCTION: TETANUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
 CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
 CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
 CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
 CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
 CC ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 76-GLN-|-PHE-77
 CC BOND OF SYNAPTOSOMAL-2.
 CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF 76-GLN-|-PHE-77 BOND IN
 CC SYNAPTOSOMAL-2.
 CC -!- SUBUNIT: THE PRECURSOR POLYPEPTIDE IS SUBSEQUENTLY CLEAVED TO
 CC YIELD SUBCHAINS L AND H. THESE REMAIN LINKED BY A DISULFIDE BRIDGE
 CC AND ARE NON-TOXIC AFTER SEPARATION.
 CC -!- THE C-TERMINAL OF THE HEAVY CHAIN BINDS TO GANGLIOSIDE RECEPTORS.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27 (ZINC
 CC METALLOPROTEASE); ALSO KNOWN AS THE TETANUS/BOTULINUM NEUROTOXIN
 CC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X04436; G40770; -.
 DR EMBL; M12739; G144921; -.
 DR EMBL; X06214; G40774; -.
 DR PIR; A25689; BTCLTN.
 DR PDB; IAF9; 29-APR-98.
 DR PROSITE; PS00142; ZINC_PROTEASE; 1.
 DR NEUROTOXIN; TRANSMEMBRANE; HYDROLASE; METALLOPROTEASE; ZINC; PLASMID;
 KW 3D-STRUCTURE.
 KW INIT_MET 0 0
 FT CHAIN 1 456 TETANUS TOXIN LIGHT CHAIN.
 FT CHAIN 457 1314 TETANUS TOXIN HEAVY CHAIN.
 FT METAL 232 232 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 233 233 BY SIMILARITY.
 FT METAL 236 236 ZINC (CATALYTIC) (BY SIMILARITY).
 FT TRANSMEM 226 246 POTENTIAL.
 FT TRANSMEM 669 689 POTENTIAL.
 FT DISULFID 438 466 INTERCHAIN.
 FT DISULFID 1076 1092
 SQ SEQUENCE 1314 AA; 150550 MW; 26190E3E CRC32;

Query Match 100.0%; Score 158; DB 1; Length 1314;
 Best Local Similarity 100.0%; Pred. No. 2.29e-20;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 946 FNNFTVSWLRVPKVSASHLE 966
 QY 1 FNNFTVSWLRVPKVSASHLE 21
 RESULT 2
 ID BXG_CLOBO STANDARD; PRT; 1296 AA.
 AC Q60393;
 DT 01-NOV-1997 (REL. 35, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE BOTULINUM NEUROTOXIN TYPE G PRECURSOR (EC 3.4.24.69) (BONT/G)
 DE (BONTXILYSIN G).
 GN BOTG.
 OS CLOSTRIDIUM BOTULINUM.
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
 OC CLOSTRIDIUM.
 [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN-113 / 30;
 RX MEDLINE; 94093745.
 RA CAMPBELL K., COLLINS M.D., EAST A.K.;
 RT "Nucleotide sequence of the gene coding for Clostridium botulinum
 RT (Clostridium argentinense) type G neurotoxin: genealogical comparison
 RT with other clostridial neurotoxins.";
 RL BIOCHIM. BIOPHYS. ACTA 1216:487-491(1993).
 CC -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
 CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
 CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
 CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
 CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
 CC ENDOPEPTIDASE.
 CC -!- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
 CC A HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL
 CC ACTIVITY, WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDATE
 CC CHANNEL FORMATION AND TOXIN BINDING, RESPECTIVELY.
 CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
 CC -!- THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF BOTULINUM
 CC NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27 (ZINC
 CC METALLOPROTEASE); ALSO KNOWN AS THE TETANUS/BOTULINUM NEUROTOXIN
 CC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X74162; G441276; -.
 DR PROSITE; PS00142; ZINC_PROTEASE; 1.
 DR PFAM; PF000099; zn-protease; 1.
 DR HSP; P04956; IAF9.
 KW NEUROTOXIN; HYDROLASE; METALLOPROTEASE; ZINC.
 KW INIT_MET 0 0
 FT CHAIN 1 441 BOTULINUM NEUROTOXIN G, LIGHT-CHAIN.
 FT CHAIN 442 1296 BOTULINUM NEUROTOXIN G, HEAVY-CHAIN.
 FT METAL 229 229 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 230 230 BY SIMILARITY.
 FT METAL 233 233 ZINC (CATALYTIC) (BY SIMILARITY).
 FT DISULFID 435 449 INTERCHAIN (PROBABLE).
 SQ SEQUENCE 1296 AA; 149013 MW; 8570B903 CRC32;

Query Match 60.1%; Score 95; DB 1; Length 1296;
 Best Local Similarity 38.1%; Pred. No. 4.20e-06;
 Matches 8; Conservative 9; Mismatches 4; Indels 0; Gaps 0;

Db 929 FDNFSINFWVRTPKYNNDIQ 949
 QY 1 FNNFTVSWLRVPKVSASHLE 21

```

DR EMBL; S73676; G765179; -.
DR EMBL; X70820; G407791; -.
DR EMBL; X70816; G407789; -.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PFAM; PF00099; zn-protease; 1.
DR HSP; P04958; IAF9.
KW NEUROTOXIN; TRANSMEMBRANE; HYDROLASE; METALLOPROTEASE; ZINC.
FT CHAIN 1 436
FT CHAIN 437 1274
FT METAL 227 227
FT ACT_SITE 228 228
FT METAL 231 231
FT DISULFID 429 445
FT SEQUENCE 1274 AA; 146709 MW; 10954D22 CRC32;
Query Match 58.9%; Score 93; DB 1; Length 1274;
Best Local Similarity 57.1%; Pred.No.1.08e-05;
Matches 8; Conservative 5; Mismatches 1; Indels 0; Gaps
Db 930 YONFISFWVRPK 943
QY 1 FNNFTVFWVRPK 14
: |||:||||:|
RESULT 4
ID BXB_CLOBO STANDARD; PRT; 1290 AA.
AC P10844; P10843;
DT 01-JUL-1989 (REL. 11, CREATED)
DI 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DI 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE BOTULINUM NEUROTOXIN TYPE B PRECURSOR (EC 3.4.24.69) (BONT/B)
DE (BONTOKILYSIN B).
OS BOTB.
OS CLOSTRIDIUM BOTULINUM.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
OC CLOSTRIDIUM.
RN [1]
RX SEQUENCE FROM N.A.
RX MEDLINE; 92384550.
RA WHELAN S.M., ELMORE M.J., BODSWORTH N.J., BREHM J.K., ATKINSON I.,
RA MINTON N.P.;
RT "Molecular cloning of the Clostridium botulinum structural gene
RT encoding the type B neurotoxin and determination of its entire
RT nucleotide sequence."
RL APPL. ENVIRON. MICROBIOL. 58:2345-2354(1992).
RN [2]
RX SEQUENCE OF 35-245 FROM N.A.
RX STRAIN-NCTC 7273;
RC SZABO E.A., PEMBERTON J.M., DESMARCHELIER P.M.;
RL SUBMITTED (APR-1992) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RX SEQUENCE OF 633-993 FROM N.A.
RX STRAIN-NCTC 7273;
RC MEDLINE; 94013372.
RA CAMPBELL K., EAST A.K., COLLINS M.D.;
RT "Gene probes for identification of the botulinum neurotoxin gene and
RT specific identification of neurotoxin types B, E, and F.;"
RL J CLIN. MICROBIOL. 31:2255-2262(1993).
RN [4]
RX SEQUENCE OF 1-44 AND 441-466.
RC STRAIN=657;
RX MEDLINE; 89000987.
RA DASGUPTA B.R., DATTA A.;
RT "Botulinum neurotoxin type B (strain 657): partial sequence and
RT similarity with tetanus toxin.;"
RL BIOCHIMIE 70:811-817(1988).
RN [5]
RX SEQUENCE OF 1-16 AND 441-458.
RC STRAIN=OKRA;
RX MEDLINE; 85197963.
RA SCHMIDT J.J., SATHYAMOORTHY V., DASGUPTA B.R.;
RT "Partial amino acid sequences of botulinum neurotoxins types B and
RT E.;"

```

RL ARCH. BIOCHEM. BIOPHYS. 238:544-548(1985).
 RN [6]
 RX IDENTIFICATION AS ZINC-PROTEASE.
 RA SCHIAVO G., ROSSETTO O., SANTUCCI A., DASGUPTA B.R., MONTECUCCO C.;
 RT "Botulinum neurotoxins are zinc proteases.";
 RL J. BIOL. CHEM. 267:23479-23483(1992).
 RN [7]
 RX IDENTIFICATION OF SUBSTRATE.
 RA SCHIAVO G., BENFENATI F., POULAIN B., ROSSETTO O., DE LAURETO P.P.,
 RA DASGUPTA B.R., MONTECUCCO C.;
 RT "Tetanus and botulinum-B neurotoxins block neurotransmitter release
 by proteolytic cleavage of synaptobrevin.";
 RL NATURE 359:832-835(1992).
 CC -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
 RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
 AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
 WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
 INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
 ENDOPEPTIDASE THAT CLEAVES THE 76-GLN-|-PHE-77 BOND OF
 SYNAPTOSOMAL-2.
 CC -!- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
 A HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL
 ACTIVITY, WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE
 CHANNEL FORMATION AND TOXIN BINDING, RESPECTIVELY.
 CC -!- SUBCELLULAR LOCATION: SECRETED.
 CC -!- THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF BOTULINUM
 NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27 (ZINC
 METALLOPROTEASE); ALSO KNOWN AS THE TETANUS/BOTULINUM NEUROTOXIN
 SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M81186; G144735; -
 DR EMBL; Z11934; G40384; -
 DR EMBL; X70817; G407783; -
 DR PIR; S07128; S07128.
 DR PIR; S07155; S07155.
 DR PIR; S08562; S08562.
 DR PIR; S08573; S08573.
 DR PIR; S08574; S08574.
 DR PIR; A48940; A48940.
 DR PROSITE; PS00142; ZINC_PROTEASE; 1.
 DR PFAM; PF00099; zn-protease; 1.
 DR HSSP; P04958; IAF9.
 KW NEUROTOXIN; TRANSMEMBRANE; HYDROLASE; METALLOPROTEASE; ZINC.
 FT INIT MET 0
 FT CHAIN 1 440 BOTULINUM NEUROTOXIN B; LIGHT-CHAIN.
 FT CHAIN 441 1290 BOTULINUM NEUROTOXIN B; HEAVY-CHAIN.
 FT METAL 229 229 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 230 230 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 233 233 ZINC (CATALYTIC) (BY SIMILARITY).
 FT DISULFID 436 445 INTERCHAIN (PROBABLE).
 FT CONFLICT 29 29 T -> M (IN REF. 4).
 FT CONFLICT 217 217 R -> G (IN REF. 2).
 FT CONFLICT 224 224 A -> S (IN REF. 2).
 FT CONFLICT 463 463 S -> R (IN REF. 4).
 SQ SEQUENCE 1290 AA; 150670 MW; 4799BEC70 CRC32;
 Query Match 58.9%; Score 93; DB 1; Length 1290;
 Best Local Similarity 64.3%; Pred. No. 1.08e-05;
 Matches 9; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
 Db 922 FLDFSVFWIRPK 935
 | :|:||||:|:|

QY 1 FNFVTFWLRVVK 14
 RESULT 5
 ID BXA_CLOBO STANDARD; PRT; 1295 AA.
 AC P10845; P18639; P01561;
 DT 01-JUL-1989 (REL. 11, CREATED)
 DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE BOTULINUM NEUROTOXIN TYPE A PRECURSOR (EC 3.4.24.69) (BONT/A)
 DE (BONTXILYSIN A).
 GN BOTA OR BNA.
 OS CLOSTRIDIUM BOTULINUM.
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
 CC CLOSTRIDIUM.
 [1]
 EN SEQUENCE FROM N.A.
 RP STRAIN-NCIC 2916;
 RC MEDLINE; 90235864.
 RA THOMPSON D.E., BREHM J.K., OULTRAM J.D., SWINFIELD T.-J.,
 RA SHONE C.C., ATKINSON T., MELLING J., MINTON N.P.;
 RT "The complete amino acid sequence of the Clostridium botulinum type A
 neurotoxin, deduced by nucleotide sequence analysis of the encoding
 gene.";
 RL EUR. J. BIOCHEM. 189:73-81(1990).
 [2]
 EN SEQUENCE FROM N.A.
 RP STRAIN-62A;
 RC MEDLINE; 90264400.
 RA BINZ B., KUARKONO H., WILLE M., PREVENT J., WERNARS K., NIEMANN H.;
 RT "The complete sequence of botulinum neurotoxin type A and comparison
 with other clostridial neurotoxins.";
 RL J. BIOL. CHEM. 265:9153-9158(1990).
 [3]
 EN SEQUENCE OF 1-34 FROM N.A.
 RP STRAIN-HALL;
 RC MEDLINE; 89350959.
 RA BETLEY M.J., SOMERS E., DASGUPTA B.R.;
 RT "Characterization of botulinum type A neurotoxin gene: delineation of
 the N-terminal encoding region.";
 RL BIOCHEM. BIOPHYS. RES. COMMUN. 162:1388-1395(1989).
 [4]
 EN SEQUENCE OF 1-16.
 RP MEDLINE; 84178501.
 RA SCHMIDT J.J., SARTYMOORTHY V., DASGUPTA B.R.;
 RT "Partial amino acid sequence of the heavy and light chains of
 botulinum neurotoxin type A.";
 RL BIOCHEM. BIOPHYS. RES. COMMUN. 119:900-904(1984).
 [5]
 EN SEQUENCE OF 1-46.
 RP DASGUPTA B.R., FOLEY J., NIECE R.;
 RT "Partial sequence of the light chain of botulinum neurotoxin type A.";
 RL BIOCHEMISTRY 26:4162-4162(1987).
 [6]
 EN SEQUENCE OF 1-5 AND 444-456.
 RP MEDLINE; 91120847.
 RA DASGUPTA B.R., DEKLEVA M.L.;
 RT "Botulinum neurotoxin type A: sequence of amino acids at the
 N-terminus and around the nicking site.";
 RL BIOCHIMIE 72:661-664(1990).
 [7]
 EN SEQUENCE OF 448-464 AND 872-895.
 RP MEDLINE; 89024662.
 RA SATHYMOORTHY V., DASGUPTA B.R., FOLEY J., NIECE R.L.;
 RT "Botulinum neurotoxin type A: cleavage of the heavy chain into two
 halves and their partial sequences.";
 RL ARCH. BIOCHEM. BIOPHYS. 266:142-151(1988).
 [8]
 EN SEQUENCE OF 448-482.
 RP MEDLINE; 85285016.
 RA SHONE C.C., HAMBLETON P., MELLING J.;
 RT "Inactivation of Clostridium botulinum type A neurotoxin by trypsin
 and purification of two tryptic fragments. Proteolytic action near

the COOH-terminus of the heavy subunit destroys toxin-binding activity.;
 EUR. J. BIOCHEM. 151:75-82(1985).
 [9]
 RP IDENTIFICATION OF SUBSTRATE.
 RX MEDLINE: 94063091.
 RA SCHIAVO G., SANTUCCI A., DASGUPTA B.R., MEHTA P.P., JONTES J., BENFENATI F., WILSON M.C., MONTECUCCI C.;
 "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct COOH-terminal peptide bonds.";
 FEBS LETT. 335:99-103(1993).
 [10]
 RP IDENTIFICATION OF SUBSTRATE.
 RX MEDLINE: 94124495.
 RA BINZ T., BLASI J., YAMASAKI S., BAUMEISTER A., LINK E., SUDHOF T.C., JAHN R., NIEMANN H.;
 "Proteolysis of SNAP-25 by types E and A botulinum neurotoxins.";
 J. BIOL. CHEM. 269:1617-1620(1994).
 CC -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC ENDOPEPTIDASE. IT CATALYZES THE HYDROLYSIS OF THE 197-GLN-|-ARG-198 BOND IN SNAP-25.
 CC -!- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY, WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL FORMATION AND TOXIN BINDING, RESPECTIVELY.
 CC -!- SUBCELLULAR LOCATION: SECRETED.
 CC -!- THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27 (ZINC METALLOPROTEASE); ALSO KNOWN AS THE TETANUS/BOTULINUM NEUROTOXIN SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 DR EMBL: X52066; G40382; -.
 DR EMBL: M30196; G144865; -.
 DR EMBL: M27892; G551776; -.
 DR PIR: A35294; BICLAB.
 DR PIR: S09492; S09492.
 DR PROSITE: PS00142; ZINC_PROTEASE; 1.
 DR PFAM: PF00099; zn-protease; 1.
 DR HSSP: P04958; IAF9.
 KW NEUROTOXIN; TRANSMEMBRANE; HYDROLASE; METALLOPROTEASE; ZINC.
 FT INIT MET 0 0
 FT CHAIN 1 447
 FT CHAIN 448 1295
 FT METAL 222 222
 FT ACT_SITE 223 223
 FT METAL 226 226
 FT METAL 429 433
 FT DISULFID 436 446
 FT TRANSMEM 626 646
 FT TRANSMEM 655 675
 FT VARIANT 26 26
 FT V -> A.
 FT P -> Q (IN REF. 1).
 FT E -> P (IN REF. 7).
 FT T -> I (IN REF. 6).
 FT S -> K (IN REF. 6).
 FT CONFLICT 891 891
 FT CONFLICT 1295 AA; 149322 MW; 2E333BFB CRC32;
 SQ SEQUENCE 1295 AA; 149322 MW; 2E333BFB CRC32;
 Query Match 57.0%; Score 90; DB 1; Length 1295;
 Best Local Similarity 57.1%; Pred. No. 4.41e-05;
 Matches 8; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 937 YENSTSFWRIPK 950
 QY 1 FNNFVFWLRVPK 14
 RESULT 6
 ID BXE_CLOBO STANDARD; PRT; 1250 AA.
 AC Q00496;
 DT 01-JUL-1993 (REL. 26, CREATED)
 DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
 DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
 DE BOTULINUM NEUROTOXIN TYPE E PRECURSOR (EC 3.4.24.69) (BONT/E)
 DE (BONTXILYSIN E).
 OS CLOSTRIDIUM BOTULINUM.
 OS BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
 OC CLOSTRIDIUM.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BEUGA;
 RX MEDLINE: 92181428.
 RA POULET S., HAUSER D., QUANZ M., NIEMANN H., POPOFF M.R.;
 "Sequences of the botulinum neurotoxin E derived from Clostridium botulinum type E (strain Beluga) and Clostridium butyricum (strains ATCC 43181 and ATCC 43755).";
 RT ATCC 43181 and ATCC 43755.";
 RL BIOCHEM. BIOPHYS. RES. COMMUN. 183:107-113(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 92174922.
 RA WHELAN S.M., ELMORE M.J., BODSWORTH N.J., ATKINSON T., MINTON N.P.;
 "The complete amino acid sequence of the Clostridium botulinum type-E neurotoxin, derived by nucleotide sequence analysis of the encoding gene.";
 RL EUR. J. BIOCHEM. 204:657-667(1992).
 RN [3]
 RP SEQUENCE OF 1-251 FROM N.A.
 RX MEDLINE: 90264400.
 RA BINZ T., KURAZONO H., WILLE M., FREVERT J., WERNARS K., NIEMANN H.;
 "The complete sequence of botulinum neurotoxin type A and comparison with other clostridial neurotoxins.";
 RL J. BIOL. CHEM. 265:9153-9158(1990).
 RN [4]
 RP SEQUENCE OF 1-13.
 RX MEDLINE: 85197963.
 RA SCHMIDT J.J., SATHYAMOORTHY V., DASGUPTA B.R.;
 "Partial amino acid sequences of botulinum neurotoxins types B and E.";
 RL ARCH. BIOCHEM. BIOPHYS. 238:544-548(1985).
 RN [5]
 RP SEQUENCE OF 419-426.
 RX MEDLINE: 90344918.
 RA GIMENEZ J.A., DASGUPTA B.R.;
 "Botulinum neurotoxin type E fragmented with endoproteinase Lys-C reveals the site trypsin nicks and homology with tetanus neurotoxin.";
 RL BIOCHIMIE 72:213-217(1990).
 RN [6]
 RP IDENTIFICATION OF SUBSTRATE.
 RX MEDLINE: 94063091.
 RA SCHIAVO G., SANTUCCI A., DASGUPTA B.R., MEHTA P.P., JONTES J., BENFENATI F., WILSON M.C., MONTECUCCI C.;
 "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct COOH-terminal peptide bonds.";
 RL FEBS LETT. 335:99-103(1993).
 RN [7]
 RP IDENTIFICATION OF SUBSTRATE.
 RX MEDLINE: 94124495.
 RA BINZ T., BLASI J., YAMASAKI S., BAUMEISTER A., LINK E., SUDHOF T.C., JAHN R., NIEMANN H.;
 "Proteolysis of SNAP-25 by types E and A botulinum neurotoxins.";
 RL J. BIOL. CHEM. 269:1617-1620(1994).
 CC -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD

CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
 CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
 CC ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 180-ARG-1-ILE-
 CC 181 BOND IN SNAP-25.
 CC -1- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
 CC A HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL
 CC ACTIVITY, WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE
 CC CHANNEL FORMATION AND TOXIN BINDING, RESPECTIVELY.
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF BOTULINUM
 CC NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27 (ZINC
 CC METALLOPROTEASE); ALSO KNOWN AS THE TETANUS/BOTULINUM NEUROTOXIN
 CC SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X62089; G40394; -;
 CC EMBL; X62683; G40398; -;
 CC PIR; A60027; A60027;
 CC PIR; B35294; B35294;
 CC PIR; JH0257; JH0257;
 CC PIR; S08575; S08575;
 CC PIR; S18111; S18111;
 CC PIR; S21178; S21178;
 CC PROSITE; PS00142; ZINC_PROTEASE; 1;
 CC PFAM; PF00099; zn-protease; 1;
 CC HSSP; P04958; 1A99.
 CC
 CC NEUROTOXIN; TRANSMEMBRANE; HYDROLASE; METALLOPROTEASE; ZINC.
 CC
 CC INIT MET 0
 CC CHAIN 1 421 BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.
 CC CHAIN 422 1250 BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.
 CC METAL 211 211 ZINC (CATALYTIC) (BY SIMILARITY).
 CC ACT_SITE 212 212 BY SIMILARITY.
 CC METAL 215 215 ZINC (CATALYTIC) (BY SIMILARITY).
 CC DISULFID 411 425 INTERCHAIN (PROBABLE).
 CC CONFLICT 176 176 R -> G (IN REF. 2).
 CC CONFLICT 197 197 C -> S (IN REF. 2 AND 3).
 CC CONFLICT 339 339 R -> A (IN REF. 2).
 CC CONFLICT 772 772 I -> L (IN REF. 2).
 CC CONFLICT 962 963 FE -> LQ (IN REF. 2).
 CC CONFLICT 966 966 R -> A (IN REF. 2).
 CC CONFLICT 1194 1194 N -> NN (IN REF. 2).
 CC SEQUENCE 1250 AA; 143712 MW; 4BE9332D CRC32;
 CC
 CC Query Match 55.7%; Score 88; DB 1; Length 1250;
 CC Best Local Similarity 50.0%; Pred.No. 1.11e-04;
 CC Matches 7; Conservative 7; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC Db 911 YKNFSISFWVRIPN 924
 CC QY 1 FNNFTVSFWLRVPEK 14
 CC
 CC RESULT 7
 CC ID BKE_CLOBU STANDARD; PRT; 1250 AA.
 CC AC P30995;
 CC DT 01-JUL-1993 (REL. 26, CREATED)
 CC DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
 CC DE 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
 CC DE BOTULINUM NEUROTOXIN TYPE E PRECURSOR (EC 3.4.24.69) (BONT/E)
 CC DE (BONTOXILYSIN E).
 CC OS CLOSTRIDIUM BUTYRICUM.
 CC OC BACTERIA: FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
 CC CLOSTRIDIUM.
 CC RN [1]
 CC RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 43181, AND ATCC 43755;
 RX MEDLINE; 92181428.
 RA POULET S., HAUSER D., QUANZ M., NIEMANN H., POPOFF M.R.;
 RT "Sequences of the botulin neurotoxin E derived from Clostridium
 RT botulinum type E (strain Beluga) and Clostridium butyricum (strains
 RT ATCC 43181 and ATCC 43755).";
 RL BIOCHEM. BIOPHYS. RES. COMMUN. 183:107-113(1992).
 RN [2]
 RP SEQUENCE OF 1-251 FROM N.A.
 RC STRAIN=BL6340;
 RX MEDLINE; 91237316.
 RA FUGII N., KIMURA K., MURAKAMI T., INDOH T., TSUZUKI K.,
 RA YOKOSAWA N., YASHIKI T., OGUMA K.;
 RT "Cloning of a DNA fragment encoding the 5'-terminus of the botulinum
 RT type E toxin gene from Clostridium butyricum strain BL6340.";
 RL J. GEN. MICROBIOL. 137:519-525(1991).
 RN [3]
 RP SEQUENCE OF 1-48.
 RC STRAIN=5262;
 RA GIMENEZ J., FOLEY J., DASGUPTA B.R.;
 RT "Neurotoxin type E from Clostridium botulinum and C. butyricum;
 RT partial sequence and comparison.";
 RL FASEB J. 2:AI750-AI750(1988).
 CC -1- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
 CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
 CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
 CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
 CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
 CC ENDOPEPTIDASE.
 CC -1- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
 CC A HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL
 CC ACTIVITY, WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE
 CC CHANNEL FORMATION AND TOXIN BINDING, RESPECTIVELY.
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF BOTULINUM
 CC NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27 (ZINC
 CC METALLOPROTEASE); ALSO KNOWN AS THE TETANUS/BOTULINUM NEUROTOXIN
 CC SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X62088; G40380; -;
 CC EMBL; X53180; G40408; -;
 CC PIR; JH0256; JH0256.
 CC PIR; S16145; S16145.
 CC PROSITE; PS00142; ZINC_PROTEASE; 1.
 CC PFAM; PF00099; zn-protease; 1.
 CC HSSP; P04958; 1A99.
 CC
 CC NEUROTOXIN; TRANSMEMBRANE; HYDROLASE; METALLOPROTEASE; ZINC.
 CC
 CC INIT MET 0
 CC CHAIN 1 421 BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.
 CC CHAIN 422 1250 BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.
 CC METAL 211 211 ZINC (CATALYTIC) (BY SIMILARITY).
 CC ACT_SITE 212 212 BY SIMILARITY.
 CC METAL 215 215 ZINC (CATALYTIC) (BY SIMILARITY).
 CC DISULFID 411 425 INTERCHAIN (PROBABLE).
 CC CONFLICT 229 229 K -> M (IN REF. 2).
 CC SEQUENCE 1250 AA; 143265 MW; BC046B39 CRC32;
 CC
 CC Query Match 55.7%; Score 88; DB 1; Length 1250;
 CC Best Local Similarity 50.0%; Pred.No. 1.11e-04;
 CC Matches 7; Conservative 7; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC Db 911 YKNFSISFWVRIPN 924
 CC QY 1 FNNFTVSFWLRVPEK 14

```

RESULT 8
ID VP2_AHSV6 STANDARD; PRT; 1051 AA.
AC O71024;
DT 15-DEC-1998 (REL. 37, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE OUTER CAPSID PROTEIN VP2.
GN S2 OR L2.
OS AFRICAN HORSE SICKNESS VIRUS 6 (AHSV-6) (AFRICAN HORSE SICKNESS VIRUS
(SEROTYPE 6)).
OC VIRUSES; DSRNA VIRUSES; REOVIRIDAE; ORBIVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98278331.
RA WILLIAMS C.F., INQUE T., LUCUS A.-M., ZANOTTO P., ROY P.;
RT "The complete sequence of four major structural proteins of African
horse sickness virus serotype 6: evolutionary relationships within
and between the orbiviruses."
RL VIRUS RES. 53:53-73(1998).
CC -!- FUNCTION: THE VP2 PROTEIN IS ONE OF THE TWO PROTEINS (WITH VP5)
WHICH CONSTITUTE THE VIRUS PARTICLE OUTER CAPSID. IT IS THE
MAJOR TARGET OF THE HOST IMMUNOGENIC RESPONSE.
CC -!- SIMILARITY: BELONGS TO THE REOVIRUSES VP2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF021235; G3168848; -
KW COAT PROTEIN.
SQ SEQUENCE 1051 AA; 122326 MW; 05DB1A29 CRC32;

Query Match 55.1%; Score 87; DB 1; Length 1051;
Best Local Similarity 47.6%; Pred. No. 1.76e-04;
Matches 10; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Db 636 FSKRFVSQYVKEIKTKHLE 656
::: ||| |||::: |||
QY 1 FNNFTVSEWLRVPKVSASHLE 21

RESULT 9
ID VNSS_ISWV1 STANDARD; PRT; 464 AA.
AC P26002;
DT 01-MAY-1992 (REL. 22, CREATED)
DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE NONSTRUCTURAL PROTEIN NS-S.
GN NS.
OS TOMATO SPOTTED WILT VIRUS (STRAIN BRAZILIAN BR-01) (TSWV).
OC VIRUSES; SSRNA NEGATIVE-STRAND VIRUSES; BUNYAVIRIDAE; TOSPOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90264829.
RA DE HAAN P., WAGENMAKERS L., PETERS D., GOLDBACH R.;
RT "The S RNA segment of tomato spotted wilt virus has an ambisense
character."
RL J. GEN. VIROL. 71:1001-1007(1990).
CC -!- FUNCTION: FORMS FILAMENTOUS INCLUSION BODIES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----

```

```

CC EMBL; D00645; D1000995; -
DR PIR; JQ0547; MNVUWC.
KW NONSTRUCTURAL PROTEIN.
SQ SEQUENCE 464 AA; 52448 MW; BC073BFC CRC32;

Query Match 53.2%; Score 84; DB 1; Length 464;
Best Local Similarity 64.3%; Pred. No. 6.88e-04;
Matches 9; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 239 SHFKLSLWLRVPKV 252
::: |||::: |||
QY 2 NNFTVSFWLRVPKV 15

RESULT 10
ID VNSS_TSWVL STANDARD; PRT; 467 AA.
AC P26003;
DT 01-MAY-1992 (REL. 22, CREATED)
DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE NONSTRUCTURAL PROTEIN NS-S.
GN NS.
OS TOMATO SPOTTED WILT VIRUS (STRAIN BULGARIAN L3) (TSWV).
OC VIRUSES; SSRNA NEGATIVE-STRAND VIRUSES; BUNYAVIRIDAE; TOSPOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91132150.
RA MAISE E., IVANOVA L., BREYEL E., ADAM G.;
RT "Cloning and sequencing of the S RNA from a Bulgarian isolate of
tomato spotted wilt virus."
RL J. GEN. VIROL. 72:461-464(1991).
CC -!- FUNCTION: FORMS FILAMENTOUS INCLUSION BODIES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D13926; G222683; -
DR PIR; JQ0954; MNVUW1.
KW NONSTRUCTURAL PROTEIN.
SQ SEQUENCE 467 AA; 52413 MW; F2EF7412 CRC32;

Query Match 53.2%; Score 84; DB 1; Length 467;
Best Local Similarity 64.3%; Pred. No. 6.88e-04;
Matches 9; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 243 SHFKLSLWLRVPKV 256
::: |||::: |||
QY 2 NNFTVSFWLRVPKV 15

RESULT 11
ID BXCL_CLOBO STANDARD; PRT; 1290 AA.
AC P18640;
DT 01-NOV-1990 (REL. 16, CREATED)
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE BOTULINUM NEUROTOXIN TYPE C1 PRECURSOR (EC 3.4.24.69) (BONT/C1)
(BONTOLYLISIN C1).
OS CLOSTRIDIUM BOTULINUM.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
RN CLOSTRIDIUM.
RP SEQUENCE FROM N.A.
RX MEDLINE; 90370487.
RA HAUSER D., EKUND M.W., KURAZONA H., BINZ I., NIEMANN H., GILL D.M.,
RA BOQUET P., POPOFF M.R.;
RT "Nucleotide sequence of Clostridium botulinum C1 neurotoxin."

```

NUCLEIC ACIDS RES. 18:4924-4924(1990).

[2]
 RN SEQUENCE FROM N.A.
 RP STRAIN-TYPE C STOCKHOLM / C-ST;
 RX MEDLINE; 91024998.
 RA KIMURA K., FUJII N., TSUZUKI K., MURAKAMI T., INDOH T.,
 RA YOKOSAWA N., TAKESHI K., SYUTO B., OGUMA K.;
 RT "The complete nucleotide sequence of the gene coding for botulinum
 type C1 toxin in the C-ST phase genome.";
 RL BIOCHEM. BIOPHYS. RES. COMMUN. 171:1304-1311(1990).

[3]
 RN SEQUENCE OF 2-25.
 RP STRAIN-TYPE C STOCKHOLM / C-ST;
 RX MEDLINE; 88153072.
 RA TSUZUKI K., YOKOSAWA N., SYUTO B., OHISHI I., FUJII N., KIMURA K.,
 RA OGUMA K.;
 RT "Establishment of a monoclonal antibody recognizing an antigenic site
 common to Clostridium botulinum type B, C1, D, and E toxins and
 tetanus toxin.";
 RL INFECT. IMMUN. 56:898-902(1988).

[4]
 RN IDENTIFICATION OF SUBSTRATE.
 RP MEDLINE; 94038966.
 RA BLASI J., CHAPMAN E.R., YAWASAKI S., BINZ T., NIEMANN H., JAHN R.;
 RT "Botulinum neurotoxin C1 blocks neurotransmitter release by means of
 cleaving HPC-1/syntaxin.";
 RL EMBO J. 12:4821-4828(1993).

CC -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
 RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
 AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
 WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
 INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
 ENDOPEPTIDASE THAT CLEAVES SYNTAXIN.
 CC -!- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
 A HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL
 ACTIVITY, WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIAE
 CHANNEL FORMATION AND TOXIN BINDING, RESPECTIVELY.
 CC -!- SUBCELLULAR LOCATION: SECRETED.
 CC -!- THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF BOTULINUM
 NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
 CC -!- BOTULINUM TYPE C1 NEUROTOXIN IS SYNTHESIZED BY C STRAIN OF
 CLOSTRIDIUM BOTULINUM WHICH CARRY THE APPROPRIATE BACTERIOPHAGE.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27 (ZINC
 METALLOPROTEASE); ALSO KNOWN AS THE TETANUS/BOTULINUM NEUROTOXIN
 SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

DR EMBL; X66433; G509275; -;
 DR EMBL; X72793; G516175; -;
 DR EMBL; X53751; G14906; -;
 DR EMBL; D90210; G217781; -;
 DR EMBL; X52389; G40390; -;
 DR PIR; S11291; S11291.
 DR PIR; A35396; A35396.
 DR PIR; A43503; A43503.
 DR PROSITE; PS00142; ZINC_PROTEASE; 1.
 DR PFAM; PF00099; zn-protease; 1.
 KW NEUROTOXIN; TRANSMEMBRANE; HYDROLASE; METALLOPROTEASE; ZINC.
 FT INIT_MET 0
 FT CHAIN 1 448 BOTULINUM NEUROTOXIN C1, LIGHT-CHAIN.
 FT CHAIN 449 1290 BOTULINUM NEUROTOXIN C1, HEAVY-CHAIN.
 FT METAL 228 228 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 229 229 BY SIMILARITY.
 FT METAL 232 232 ZINC (CATALYTIC) (BY SIMILARITY).
 FT DISULFID 436 452 INTERCHAIN (PROBABLE).
 FT CONFLICT 64 84 P -> T (IN REF. 2).

SQ SEQUENCE 1290 AA; 148734 MW; D0BCB190 CRC32;
 Query Match 51.3%; Score 81; DB 1; Length 1290;
 Best Local Similarity 42.9%; Pred. No. 2.63e-03;
 Matches 6; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

Db 934 YESFSISFWIRINK 947
 QY 1 FNNFTVSWLRVPK 14
 ::::|::|::|
 :|::|::|::|
 RESULT 12
 ID VNSS_INSVN STANDARD; PRT; 449 AA.
 AC Q01811;
 DT 01-JUL-1993 (REL. 26, CREATED)
 DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
 DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
 DE NONSTRUCTURAL PROTEIN NS-S.
 GN NSS.
 OS IMPATIENS NECROTIC SPOT VIRUS (STRAIN NL-07) (NSV).
 OC VIRUSES; SSRNA NEGATIVE-STRAND VIRUSES; BUNYAVIRIDAE; TOSPOVIRUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 92331780.
 RA DE HAAN P., DE AVILA A.C., KORMELINK R., WESTERBROEK A.,
 RA GIELEN J.J., PETERS D., GOLDBACH R.;
 RT "The nucleotide sequence of the S RNA of Impatiens necrotic spot
 virus, a novel tospovirus.";
 RL FEBS LETT. 306:27-32(1992).
 CC -!- FUNCTION: FORMS FILAMENTOUS INCLUSION BODIES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

DR EMBL; X66972; G60489; -;
 DR PIR; S23158; S23158.
 KW NONSTRUCTURAL PROTEIN.
 SQ SEQUENCE 449 AA; 51197 MW; 6A87666F CRC32;
 Query Match 48.7%; Score 77; DB 1; Length 449;
 Best Local Similarity 50.0%; Pred. No. 1.51e-02;
 Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 244 FKSLWMRIPKI 255
 QY 4 FTVSEWLRVPKV 15
 |::|::|::|
 :|::|::|::|
 RESULT 13
 ID BXCN_CLOBO STANDARD; PRT; 1196 AA.
 AC P46081;
 DT 01-NOV-1995 (REL. 32, CREATED)
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
 DE BOTULINUM NEUROTOXIN TYPE C1, NONTOXIC COMPONENT.
 OS CLOSTRIDIUM BOTULINUM.
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
 CC CLOSTRIDIUM.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN-TYPE C STOCKHOLM / C-ST;
 RX MEDLINE; 92231894.
 RA TSUZUKI K., KIMURA K., FUJII N., YOKOSAWA N., OGUMA K.;
 RT "The complete nucleotide sequence of the gene coding for the
 non-toxic-nonhemagglutinin component of Clostridium botulinum type C
 progenitor toxin.";
 RL BIOCHEM. BIOPHYS. RES. COMMUN. 183:1273-1279(1992).
 CC -!- FUNCTION: THE NONTXIC COMPONENT IS NECESSARY TO MAINTAIN

```

CC -----
CC TOXICITY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; X52389; G40389; -
CC KW NEUROTOXIN.
CC SQ SEQUENCE 1196 AA; 138740 MW; E5CL1933 CRC32;
CC -----
CC Query Match 48.1%; Score 76; DB 1; Length 1196;
CC Best Local Similarity 80.0%; Pred. No. 2.32e-02;
CC Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
CC -----
CC Db 899 NNFTICFWLR 908
CC QY 2 NNFTVFWLR 11
CC -----
CC RESULT 14
CC ID BXD_CLOBO STANDARD; PRT; 1276 AA.
CC AC P19321;
CC DT 01-NOV-1990 (REL. 16, CREATED)
CC DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
CC DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
CC DE BOTULINUM NEUROTOXIN TYPE D PRECURSOR (EC 3.4.24.69) (BONT/D)
CC DE (BONTXILYSIN D).
CC GN BOTD
CC OS CLOSTRIDIUM BOTULINUM.
CC OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
CC CLOSTRIDIUM.
CC [1]
CC RN SEQUENCE FROM N.A.
CC RC STRAIN=BVD/-3;
CC RX MEDLINE; 91016853.
CC RA BINZ T., KURAZONO H., POPOFF M.R., EKLUND M.W., SAKAGUCHI G.,
CC KOZAKI S., KRIEGLSTEIN K., HENSCHEN A., GILL D.M., NIEMANN H.;
CC "Nucleotide sequence of the gene encoding Clostridium botulinum
CC neurotoxin type D."
CC RL NUCLEIC ACIDS RES. 18:5556-5556(1990).
CC [2]
CC RN SEQUENCE FROM N.A.
CC RC STRAIN=CB16;
CC RX MEDLINE; 93042276.
CC RA SUNAGAWA H., OHYAMA T., WATANABE T., INOUE K.;
CC "The complete amino acid sequence of the Clostridium botulinum type D
CC neurotoxin, deduced by nucleotide sequence analysis of the encoding
CC Phase d-16 phi genome."
CC RL J. VET. MED. SCI. 54:905-913(1992).
CC [3]
CC RN PARTIAL SEQUENCE.
CC RC STRAIN=D-SA. AND-D-1873;
CC RX MEDLINE; 89339741.
CC RA MORIISHI K., SYUTO B., KUBO S., OGUMA K.;
CC "Molecular diversity of neurotoxins from Clostridium botulinum type D
CC strains."
CC RL INFECT. IMMUN. 57:2886-2891(1989).
CC [4]
CC RN IDENTIFICATION OF SUBSTRATE.
CC RX MEDLINE; 94230352.
CC RA YAMASAKI S., BAUMEISTER A., BINZ T., BLASI J., LINK E., CORNILLE F.,
CC ROQUES B., FYKSE E.M., SUEHOF T.C., JAHN R., NIEMANN H.;
CC "Cleavage of members of the synaptobrevin/VAMP family by types D and
CC F botulinum neurotoxins and tetanus toxin."
CC RL J. BIOL. CHEM. 269:12764-12772(1994).
CC -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT

```

```

CC -----
CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC ENDOPEPTIDASE THAT CLEAVES THE 60-LYS-1-LEU-61 BOND OF
CC SYNAPTOBREVIN-1 AND -2.
CC -!- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
CC A HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL
CC ACTIVITY, WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE
CC CHANNEL FORMATION AND TOXIN BINDING, RESPECTIVELY.
CC -!- SUBCELLULAR LOCATION: SECRETED.
CC -!- THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF BOTULINUM
CC NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
CC -!- BOTULINUM TYPE D NEUROTOXIN IS SYNTHESIZED BY D STRAIN OF
CC CLOSTRIDIUM BOTULINUM WHICH CARRY THE APPROPRIATE BACTERIOPHAGE.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27 (ZINC
CC METALLOPROTEASE); ALSO KNOWN AS THE TETANUS/BOTULINUM NEUROTOXIN
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X54254; G40396; -
CC EMBL; S49407; G260239; -
CC PIR; S11455; S11455.
CC PROSITE; PS00142; ZINC_PROTEASE; 1.
CC PFAM; PF00099; zn-protease; 1.
CC KW NEUROTOXIN; TRANSMEMBRANE; HYDROLASE; METALLOPROTEASE; ZINC.
CC FT CHAIN 1 442 BOTULINUM NEUROTOXIN D, LIGHT-CHAIN.
CC FT CHAIN 443 1276 BOTULINUM NEUROTOXIN D, HEAVY-CHAIN.
CC FT METAL 229 229 ZINC (CATALYTIC) (BY SIMILARITY).
CC FT ACT_SITE 230 230 BY SIMILARITY.
CC FT METAL 233 233 ZINC (CATALYTIC) (BY SIMILARITY).
CC FT DISULFID 437 450 INTERCHAIN (PROBABLE).
CC FT VARIANT 15 16 ND -> PV (IN STRAIN D-SA).
CC FT VARIANT 17 18 ND -> LQ (IN STRAIN D-1873).
CC FT VARIANT 452 452 K -> Q (IN STRAIN D-SA).
CC FT VARIANT 457 457 R -> T (IN STRAIN D-SA).
CC FT VARIANT 457 457 R -> F (IN STRAIN D-1873).
CC FT VARIANT 462 462 A -> D (IN STRAIN D-1873).
CC FT VARIANT 489 489 K -> N (IN STRAIN CB16).
CC FT VARIANT 644 644 N -> K (IN STRAIN CB16).
CC FT VARIANT 1122 1122 Q -> R (IN STRAIN CB16).
CC SQ SEQUENCE 1276 AA; 146871 MW; 02D7EECC CRC32;
CC -----
CC Query Match 45.6%; Score 72; DB 1; Length 1276;
CC Best Local Similarity 40.9%; Pred. No. 1.25e-01;
CC Matches 9; Conservative 9; Mismatches 3; Indels 1; Gaps 1;
CC -----
CC Db 928 YENSSVFWIKISKDLTNSHNE 949
CC QY 1 FNNFTVFWLRVPK-VSASHLE 21
CC -----
CC RESULT 15
CC ID BXEN_CLOBO STANDARD; PRT; 1162 AA.
CC AC P46082;
CC DT 01-NOV-1995 (REL. 32, CREATED)
CC DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
CC DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
CC DE BOTULINUM NEUROTOXIN TYPE E, NONTXIC COMPONENT.
CC GN ENT-120.
CC OS CLOSTRIDIUM BOTULINUM.
CC OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
CC CLOSTRIDIUM.
CC [1]
CC RN SEQUENCE FROM N.A.
CC RC STRAIN=MASHIRE;
CC RX MEDLINE; 93195515.
CC RA FUJII N., KIMURA K., YOKOSAWA N., YASHIKI I., TSUZUKI K., OGUMA K.;
CC "The complete nucleotide sequence of the gene encoding the nontoxic

```

Tue Aug 17 16:13:10 1999

```

RT component of Clostridium botulinum type E progenitor toxin."
RL J. GEN. MICROBIOL. 139:79-86(1993).
CC -!- FUNCTION: THE NONTXIC COMPONENT IS NECESSARY TO MAINTAIN
CC TOXICITY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D12697; G285708; -.
KW NEUROTOXIN.
SQ SEQUENCE 1162 AA; 136856 MW; F650831D CRC32;

Query Match 44.3%; Score 70; DB 1; Length 1162;
Best Local Similarity 70.0%; Pred. No. 2.84e-01;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 864 NNFSIYFWLR 873
QY 2 NNFTVSEWLR 11

Search completed: Tue Aug 17 15:54:43 1999
Job time : 15 secs.

```

W P E R E H
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Tue Aug 17 15:55:01 1999; MasPar time 6.25 Seconds
Tabular output not generated. 183.290 Million cell updates/sec

Title: >US-09-049-847-2
Description: (1-21) from US09049847.pep
Perfect Score: 158
Sequence: 1 FNNFTVSFWLRVPKVSASHLE 21

Scoring table: PAM 150
Gap 15

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: spiremb19
1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human
5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organelle
9:sp-phage 10:sp-plant 11:sp-rodent 12:sp-unclassified
13:sp-vertebrate 14:sp-virus

Statistics: Mean 29.394; Variance 45.917; scale 0.640

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					
Result No.	Score	Query Match	Length	ID	Description
1	95	60.1	1268	2	NEUROTOXIN TYPE F.
2	93	58.9	361	2	BOTULINUM NEUROTOXIN T
3	93	58.9	361	2	BOTULINUM NEUROTOXIN T
4	93	58.9	1278	2	BONT/F (BOTULINUM NEUR
5	93	58.9	1291	2	BOTULINUM NEUROTOXIN T
6	88	55.7	367	2	BOTULINUM NEUROTOXIN T
7	88	55.7	367	2	BOTULINUM NEUROTOXIN T
8	85	53.8	1296	2	BOTULINUM NEUROTOXIN T
9	84	53.2	467	14	NON-STRUCTURAL PROTEIN
10	84	53.2	467	14	NON-STRUCTURAL PROTEIN
11	82	51.9	457	14	NON-STRUCTURAL PROTEIN
12	81	51.3	1285	2	NEUROTOXIN CONSISTING
13	79	50.0	322	5	COSMID 22239.
14	76	48.1	1194	2	138KDA PROTEIN ASSOCIA
15	76	48.1	1196	2	PROGENITOR TOXIN L NON
16	76	48.1	1196	9	ANTP-139.
17	75	47.5	175	5	COSMID 2C239.
18	75	47.5	558	5	PENTAXIN-LIKE PROTEIN.
19	72	45.6	406	2	HYPOTHETICAL 44.2 KD P
20	72	45.6	1280	2	NEUROTOXIN.

21	69	43.7	637	4	O14744	SKBLHS.	1.05e+00
22	69	43.7	1197	2	O33871	NONTOXIC-NONHEMAGGLUTI	1.05e+00
23	69	43.7	1197	2	O45888	NONTOXIC-NONHEMAGGLUTI	1.05e+00
24	69	43.7	1197	2	O71117	PROGENITOR TOXIN COMPL	1.05e+00
25	69	43.7	1198	2	O06018	NUCLEOLAR PHOSPHOPROTE	1.05e+00
26	68	43.0	699	4	O14978	ORF (FRAGMENT).	1.55e+00
27	68	43.0	707	4	O15030	FROM BASES 1895340 TO	1.55e+00
28	67	42.4	152	2	P76263	HYPOTHETICAL 28.0 KD P	2.28e+00
29	66	41.8	244	2	O05883	NONTOXIC-NONHEMAGGLUTI	3.33e+00
30	66	41.8	1159	2	O71109	NTNH PROTEIN.	3.33e+00
31	66	41.8	1160	2	O87710	NTNH PROTEIN.	3.33e+00
32	66	41.8	1161	2	O45891	NONTOXIC-NONHEMAGGLUT	3.33e+00
33	66	41.8	1161	2	O69276	BOTULINUM NEUROTOXIN T	3.33e+00
34	66	41.8	1165	2	O45887	NEUROTOXIN COMPLEX M N	3.33e+00
35	66	41.8	1165	2	O45844	TYPE A PROGENITOR TOXIN	3.33e+00
36	66	41.8	1193	2	O45914	A NTNH AND BONT GENES.	3.33e+00
37	66	41.8	1193	2	P71107	NONTOXIC-NONHEMAGGLUTI	3.33e+00
38	66	41.8	1193	2	P71108	NTNHA.	3.33e+00
39	66	41.8	1193	2	O45880	NTNH PROTEIN.	3.33e+00
40	66	41.8	1198	2	O45893	NONTOXIC-NONHEMAGGLUT	3.33e+00
41	66	41.8	1198	2	O69277	REPRESSOR PROTEIN OF S	4.87e+00
42	65	41.1	203	2	Q06370	BETA-MANNOSIDASE.	4.87e+00
43	65	41.1	510	1	Q51733	INTRON ORF.	4.87e+00
44	65	41.1	786	8	Q34832	NONSTRUCTURAL POLYPROT	4.87e+00
45	65	41.1	834	14	O41921		

ALIGNMENTS

RESULT 1		PRELIMINARY;		PRT; 1268 AA.	
ID	Q45851				
AC	Q45851;				
DT	01-NOV-1996 (TREMELREL. 01, CREATED)				
DT	01-NOV-1996 (TREMELREL. 01, LAST SEQUENCE UPDATE)				
DT	01-NOV-1998 (TREMELREL. 08, LAST ANNOTATION UPDATE)				
DE	NEUROTOXIN TYPE F.				
GN	BONT /F.				
OS	CLOSTRIDIUM BARATI.				
OC	BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;				
OC	CLOSTRIDIUM.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE; 93252228.				
RA	THOMPSON D.E., HUPSON R.A., EAST A.K., ALLANWAY D., COLLINS M.D.,				
RA	RICHARDSON P.T.;				
RT	"Nucleotide sequence of the gene coding for Clostridium barati type F				
RT	neurotoxin: comparison with other clostridial neurotoxins.";				
RL	FEMS MICROBIOL. LETT. 108:175-182(1993).				
DR	EMBL; X68262; G49139; -				
DR	PFAM; PF00099; zn-protease; 1.				
KW	NEUROTOXIN.				
SQ	SEQUENCE 1268 AA; 145512 MW; 37B/AA07 CRC32;				
Query Match 60.1%; Score 95; DB 2; Length 1268;					
Best Local Similarity 64.3%; Pred. No. 1.58e-05;					
Matches 9; Conservative 4; Mismatches 1; Indels 0; Gaps 0;					
Db	922 YONFSVSWFWRIPK 935				
QY	: : :				
	1 FNNFTVSFWLRVPEK 14				
RESULT 2					
ID	Q45848				
AC	Q45848;				
DT	01-NOV-1996 (TREMELREL. 01, CREATED)				
DT	01-NOV-1996 (TREMELREL. 01, LAST SEQUENCE UPDATE)				
DT	01-NOV-1998 (TREMELREL. 08, LAST ANNOTATION UPDATE)				
DE	BOTULINUM NEUROTOXIN TYPE B (FRAGMENT).				
GN	BONT/B.				
OS	CLOSTRIDIUM BOTULINUM.				
OC	BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;				
OC	CLOSTRIDIUM.				

```

RN  SEQUENCE FROM N.A.
RP  STRAIN-TYPE B, NON-PROTEOLYTIC EKLUND 2B (COLWORTH 229);
RX  MEDLINE; 94013372.
RA  CAMPBELL K., EAST A.K., COLLINS M.D.;
RT  "Gene probes for identification of the botulin neurotoxin gene and
RT  specific identification of neurotoxin types B, E, and F.";
RL  J. CLIN. MICROBIOL. 31:2255-2262(1993).
DR  EMBL; X70819; G407781; -.
KW  NEUROTOXIN.
FT  NON_TER 361 1
FT  NON_TER 361 361
SQ  SEQUENCE 361 AA; 42131 MW; 8FBF998C CRC32;

Query Match 58.9%; Score 93; DB 2; Length 361;
Best Local Similarity 64.3%; Pred. No. 3.93e-05;
Matches 9; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 290 FLDFSVSWIRPK 303
QY 1 FNNFTVSWLRVPK 14

RESULT 3
ID Q45846 PRELIMINARY; PRT; 361 AA.
AC Q45846;
DT 01-NOV-1996 (TREMELREL. 01, CREATED)
DT 01-NOV-1996 (TREMELREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMELREL. 08, LAST ANNOTATION UPDATE)
DE BOTULINUM NEUROTOXIN TYPE B (FRAGMENT).
GN BONT/B.
OS CLOSTRIDIUM BOTULINUM.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
OC CLOSTRIDIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TYPE B, NON-PROTEOLYTIC 2129B (SCOTT);
RX MEDLINE; 94013372.
RA CAMPBELL K., EAST A.K., COLLINS M.D.;
RT "Gene probes for identification of the botulin neurotoxin gene and
RT specific identification of neurotoxin types B, E, and F.";
RL J. CLIN. MICROBIOL. 31:2255-2262(1993).
DR EMBL; X70814; G40779; -.
KW NEUROTOXIN.
FT NON_TER 1 1
FT NON_TER 361 361
SQ SEQUENCE 361 AA; 42175 MW; 3863DE86 CRC32;

Query Match 58.9%; Score 93; DB 2; Length 361;
Best Local Similarity 64.3%; Pred. No. 3.93e-05;
Matches 9; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 290 FLDFSVSWIRPK 303
QY 1 FNNFTVSWLRVPK 14

RESULT 4
ID Q57236 PRELIMINARY; PRT; 1278 AA.
AC Q57236; Q45863;
DT 01-NOV-1996 (TREMELREL. 01, CREATED)
DT 01-NOV-1996 (TREMELREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMELREL. 08, LAST ANNOTATION UPDATE)
DE BONT/F (BOTULINUM NEUROTOXIN TYPE F).
GN BONT/F.
OS CLOSTRIDIUM BOTULINUM.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
OC CLOSTRIDIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NCIC 10281;
RX HUTSON R.A., COLLINS M.D.;
RA SUBMITTED (AUG-1995) TO EMBL/GENBANK/DBJ DATA BANKS.

```

```

RN  SEQUENCE FROM N.A.
RP  ELMORE M.J., BODSWORTH N.J., WHELAN S.M., MINTON N.P.;
RL  SUBMITTED (AUG-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
[3]
RN  SEQUENCE OF 535-1000 FROM N.A.
RP  STRAIN-TYPE F, PROTEOLYTIC F LANCELAND (NCIC 1028);
RX  MEDLINE; 94013372.
RA  CAMPBELL K., EAST A.K., COLLINS M.D.;
RT  "Gene probes for identification of the botulin neurotoxin gene and
RT  specific identification of neurotoxin types B, E, and F.";
RL  J. CLIN. MICROBIOL. 31:2255-2262(1993).
DR  EMBL; X81714; G971349; -.
DR  EMBL; L35496; G529984; -.
DR  EMBL; X70821; G407793; -.
DR  PFAM; PF00099; zn-protease; 1.
KW  NEUROTOXIN.
SQ  SEQUENCE 1278 AA; 147073 MW; 56C58E6F CRC32;

Query Match 58.9%; Score 93; DB 2; Length 1278;
Best Local Similarity 57.1%; Pred. No. 3.93e-05;
Matches 8; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 931 YQNFSSISWVRPK 944
QY 1 FNNFTVSWLRVPK 14

RESULT 5
ID Q08077 PRELIMINARY; PRT; 1291 AA.
AC Q08077;
DT 01-NOV-1996 (TREMELREL. 01, CREATED)
DT 01-NOV-1996 (TREMELREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMELREL. 08, LAST ANNOTATION UPDATE)
DE BOTULINUM NEUROTOXIN TYPE B (EC 3.4.24.-) (BONT/B).
GN BONT/B.
OS CLOSTRIDIUM BOTULINUM.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
OC CLOSTRIDIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-EKLUND 17B ATCC25765;
RX MEDLINE; 94122659.
RA HUTSON R.A., COLLINS M.D., EAST A.K., THOMPSON D.E.;
RT "Nucleotide sequence of the gene coding for non-proteolytic
RT clostridium botulinum type B neurotoxin: comparison with other
RT clostridial neurotoxins.";
RL CURR. MICROBIOL. 28:101-110(1994).
CC -|- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC RELEASE.
CC IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED AND
CC MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS.
CC IT INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC ENDOPEPTIDASE THAT CLEAVES SYNAPTOSOMAL-2.
CC -|- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN
CC (L) AND A HEAVY CHAIN (H).
CC THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY, WHILE THE N-
CC AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL FORMATION AND
CC TOXIN BINDING, RESPECTIVELY.
CC -|- SUBCELLULAR LOCATION: SECRETED.
CC -|- THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF BOTULINUM
CC NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
CC -|- SIMILARITY: HIGH WITH OTHER BOTULINUM NEUROTOXINS AND WITH
CC TETANUS NEUROTOXIN.
CC -|- SIMILARITY: TO OTHER ZINC METALLOPROTEINASES IN THE ACTIVE SITE
CC REGION.
DR  EMBL; X71343; G296149; -.
DR  PFAM; PF00099; zn-protease; 1.
KW  NEUROTOXIN; TRANSMEMBRANE; HYDROLASE; METALLOPROTEASE; ZINC.
SQ  SEQUENCE 1291 AA; 150513 MW; 5210A9B7 CRC32;

Query Match 58.9%; Score 93; DB 2; Length 1291;

```

QY 1 FNNFTVSEWLRVPK 14

RESULT 8

ID Q45894 PRELIMINARY; PRT; 1296 AA.

AC Q45894; P77780;

DT 01-NOV-1996 (TREMBLREL. 01, CREATED)

DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)

DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)

DE BOTULINUM NEUROTOXIN TYPE A.

GN BONT.

OS CLOSTRIDIUM BOTULINUM.

OC BACTERIA; FRMUTICES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;

OC CLOSTRIDIUM.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=KYOTO-F;

RX MEDLINE: 94143603.

RA WILLEMS A., EAST A.K., LAWSON P.A., COLLINS M.D.;

RT "Sequence of the gene coding for the neurotoxin of Clostridium

RT botulinum type A associated with infant botulism: comparison with

RT other clostridial neurotoxins.;"

RL RES. MICROBIOL. 144:547-556(1993).

RN [2]

RP SEQUENCE OF 1-65 FROM N.A.

RC STRAIN=62A;

RX MEDLINE: 97016817.

RA EAST A.K., BHANDARI M., STACEY J.M., CAMPBELL K.D., COLLINS M.D.;

RT "Organization and phylogenetic interrelationships of genes encoding

RT components of the botulinum toxin complex in proteolytic Clostridium

RT botulinum types A, B, and F: evidence of chimeric sequences in the

RT gene encoding the nontoxic nonhemagglutinin component.;"

RL INT. J. SYST. BACTERIOL. 46:1105-1112(1996).

DR EMBL: X73423; G507071; -

DR EMBL: X92973; E212189; -

DR EMBL: X87974; E184381; -

KW NEUROTOXIN.

QY SEQUENCE 1296 AA; 149410 MW; 24AF86B1 CRC32;

Query Match 53.8%; Score 85; DB 2; Length 1295;

Best Local Similarity 50.0%; Pred. No. 1.38e-03;

Matches 7; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

Db 938 YENFSTSEWIKIPK 951

QY 1 FNNFTVSEWLRVPK 14

RESULT 9

ID Q37369 PRELIMINARY; PRT; 467 AA.

AC Q37369;

DT 01-JAN-1998 (TREMBLREL. 05, CREATED)

DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)

DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)

DE NON-STRUCTURAL PROTEIN.

OS NSS.

OS TOMATO SPOTTED WILT VIRUS (TSWV).

OC VIRUSES; SSRNA NEGATIVE-STRAND VIRUSES; BUNYAVIRIDAE; TOSPOVIRUS.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=TSWV-D;

RX MEDLINE: 98240946.

RA QIU W.-P., GESKE S.M., HICKEY C.M., MOYER J.W.;

RT "Tomato spotted wilt Tospovirus genome reassortment and genome

RT segment-specific adaptation.;"

RL VIROLOGY 244:186-194(1998).

DR EMBL: AF020660; G2425149; -

QY SEQUENCE 467 AA; 52587 MW; 1D153724 CRC32;

Query Match 53.2%; Score 84; DB 14; Length 467;

Best Local Similarity 64.3%; Pred. No. 2.13e-03;

Matches 9; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 243 SHEKLSLWLRVPKV 256
 QY 2 NNETVSFWLRVPKV 15

RESULT 10
 ID Q37367 PRELIMINARY; PRT; 467 AA.

AC Q37367;
 DT 01-JAN-1998 (TREMELREL. 05, CREATED)
 DT 01-JAN-1998 (TREMELREL. 05, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMELREL. 08, LAST ANNOTATION UPDATE)
 DE NON-STRUCTURAL PROTEIN.
 GN NSS.

OS TOMATO SPOTTED WILT VIRUS (TSWV).

OC VIRUSES; SSRNA NEGATIVE-STRAND VIRUSES; BUNYAVIRIDAE; TOSPOVIRUS.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=TSWV-10;

RX MEDLINE: 98240946.

RA QIU W.-P., GESKE S.M., HICKEY C.M., MOYER J.W.;

RT "tomato spotted wilt tospovirus genome reassortment and genome

segment-specific adaptation."

RL VIROLOGY 244:186-194(1998).

DR EMBL; AF020659; G2425152; -.

SQ SEQUENCE 467 AA; 52505 MW; B3A4243F CRC32;

Query Match 53.2%; Score 84; DB 14; Length 467;

Best Local Similarity 64.3%; Pred. No. 2.13e-03;

Matches 9; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 243 SHEKLSLWLRVPKV 256
 QY 2 NNETVSFWLRVPKV 15

RESULT 11
 ID Q88900 PRELIMINARY; PRT; 467 AA.

AC Q88900;
 DT 01-NOV-1996 (TREMELREL. 01, CREATED)

DT 01-NOV-1996 (TREMELREL. 01, LAST SEQUENCE UPDATE)

DT 01-NOV-1996 (TREMELREL. 01, LAST ANNOTATION UPDATE)

DE NONSTRUCTURAL PROTEIN.

OS UNKNOWN.

OC VIRUSES; SSRNA NEGATIVE-STRAND VIRUSES; BUNYAVIRIDAE.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=TSWV-B;

RA PANG S.Z., SLIGHTOM J.L., GONSALVES D.;

RL PHYTOPATHOLOGY 83:728-733(1993).

DR EMBL; L12048; G335275; -.

KW NONSTRUCTURAL PROTEIN

SQ SEQUENCE 467 AA; 52566 MW; BD32D7C7 CRC32;

Query Match 51.9%; Score 82; DB 14; Length 467;

Best Local Similarity 57.1%; Pred. No. 5.03e-03;

Matches 8; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 243 SHEKLSLWLRVPKV 256
 QY 2 NNETVSFWLRVPKV 15

RESULT 12
 ID Q45967 PRELIMINARY; PRT; 1285 AA.

AC Q45967; 1996 (TREMELREL. 01, CREATED)

DT 01-NOV-1996 (TREMELREL. 01, LAST SEQUENCE UPDATE)

DT 01-JAN-1999 (TREMELREL. 09, LAST ANNOTATION UPDATE)

DE NEUROTOXIN CONSISTING OF BOTULINUM NEUROTOXIN D AND C1.

OS CLOSTRIDIUM BOTULINUM.

OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;

OC CLOSTRIDIUM.

RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=TYPE D, SOUTH AFRICAN;
 RX MEDLINE: 96283801.
 RA MORIISHI K., KOURA M., ABE N., FUJII N., FUJINAGA Y., INOUE K.,
 RA OGUMAD K.;
 RT "Mosaic structures of neurotoxins produced from Clostridium botulinum
 RT types C and D organisms."
 RL BIOCHIM. BIOPHYS. ACTA 1307:123-126(1996).
 DR EMBL; D38442; D1008057; -.
 DR PFAM; PF00099; zn-protease; 1.
 KW NEUROTOXIN.
 SQ SEQUENCE 1285 AA; 147366 MW; 381133E1 CRC32;

Query Match 51.3%; Score 81; DB 2; Length 1285;

Best Local Similarity 42.9%; Pred. No. 7.72e-03;

Matches 6; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

Db 931 YESFISFWIRINK 944
 QY 1 FNNFTVSFWLRVPK 14

RESULT 13
 ID P91551 PRELIMINARY; PRT; 322 AA.

AC P91551;
 DT 01-MAY-1997 (TREMELREL. 03, CREATED)

DT 01-MAY-1997 (TREMELREL. 03, LAST SEQUENCE UPDATE)

DT 01-NOV-1998 (TREMELREL. 08, LAST ANNOTATION UPDATE)

DE COSMID ZC239.

GN ZC239.8.

OS CAENORHABDITIS ELEGANS.

OC EUKARYOTA; METAZOA; NEMATODA; SECRETENTERA; RHABDITIDA;

OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RX MEDLINE: 94150718.

RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,

RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,

RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,

RA GARDNER A., GREEN P., HAWKINS T., HILLIER N., JIER M., JOHNSTON L.,

RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LAITELLE P.,

RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,

RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,

RA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,

RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,

RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.,

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

elegans."

RL NATURE 368:32-38(1994).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RA XU X., KRAMER J.;

RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RA WATERSTON R.;

RL SUBMITTED (DEC-1996) TO EMBL/GENBANK/DBJ DATA BANKS.

DR EMBL; U80842; G1707209; -.

SQ SEQUENCE 322 AA; 36968 MW; B4E0F90C CRC32;

Query Match 50.0%; Score 79; DB 5; Length 322;

Best Local Similarity 70.0%; Pred. No. 1.80e-02;

Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 4 DFTVFWLRI 13

QY 3 NFTVSFWLRV 12

RESULT 14

ID Q45916 PRELIMINARY; PRT: 1194 AA.
 AC Q45916;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
 DE 138KDA PROTEIN ASSOCIATED WITH BONT /CI-HAEMAGGLUTININ COMPLEX.
 GN CHN-138.
 OS CLOSTRIDIUM BOTULINUM.
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
 OC CLOSTRIDIUM.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA HAUSER D.F., EKUND M.W., POPOFF M.R.;
 RL SUBMITTED (MAY-1992) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; X66433; G509274; -.
 SQ SEQUENCE 1194 AA; 138509 MW; 0192832A CRC32;

Query Match 48.1%; Score 76; DB 2; Length 1194;
 Best Local Similarity 80.0%; Pred. No. 6.29e-02;
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 899 NNETICFWLR 908
 QY 2 NNETVSEWLR 11
 ||||: ||||

RESULT 15
 ID Q53550 PRELIMINARY; PRT: 1196 AA.
 AC Q53550;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE PROGENITOR TOXIN L NONTXIC-NONHEMAGGLUTININ COMPONENT (FRAGMENT).
 OS CLOSTRIDIUM BOTULINUM.
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
 OC CLOSTRIDIUM.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 96025415.
 RA OHYAMA T., WATANABE T., FUJINAGA Y., INOUE K., SUNAGAWA H., FUJII N.,
 RA INOUE K., OGUMA K.;
 RT "Characterization of nontoxic-nonhemagglutinin component of the two
 RT types of progenitor toxin (M and L) produced by Clostridium botulinum
 RT type D CB-16.";
 RL MICROBIOL. IMMUNOL. 39:457-465(1995).
 DR EMBL; S80809; G1311691; -.
 FT NON_TER 1196 1196
 SQ SEQUENCE 1196 AA; 138717 MW; B534D1B0 CRC32;

Query Match 48.1%; Score 76; DB 2; Length 1196;
 Best Local Similarity 80.0%; Pred. No. 6.29e-02;
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 899 NNETICFWLR 908
 QY 2 NNETVSEWLR 11
 ||||: ||||

Search completed: Tue Aug 17 15:55:45 1999
 Job time : 45 secs.

This Page Blank (uspto)

WQSRH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Aug 17 16:18:07 1999; MasPar time 6.95 Seconds
Tabular output not generated. 152.315 Million cell updates/sec

Title: >US-09-049-847-2
Description: (1-21) from US09049847.pap
Perfect Score: 158
Sequence: 1 FNNFTYSFWLVRPKVSASHLE 21

Scoring table: PAM 150
Gap 15

Searched: 411786 seqs, 50406085 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-pending
1:P9 2:U60 3:U7 4:U80 5:U81 6:U82 7:U83 8:U84A 9:U84B
10:U85 11:U86 12:U87 13:U88 14:U89 15:U90 16:U91 17:U92
18:U93 19:NEWP 20:NEWU6 21:NEWU8 22:NEWU9

Statistics: Mean 22.277; Variance 74.265; scale 0.300

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	158	100.0	21	8	US-08-432-Sequence 3, Applicatio	1.00e-08
2	158	100.0	21	14	US-08-945-Sequence 3, Applicatio	1.00e-08
3	158	100.0	21	3	US-07-678-Sequence 1, Applicatio	1.00e-08
4	158	100.0	21	15	US-09-049-Sequence 2, Applicatio	1.00e-08
5	158	100.0	21	16	US-09-171-Sequence 10, Applicati	1.00e-08
6	158	100.0	21	12	US-08-724-Sequence 5, Applicatio	1.00e-08
7	158	100.0	21	8	US-08-406-Sequence 4, Applicatio	1.00e-08
8	158	100.0	21	5	US-08-161-Sequence 56, Applicatio	1.00e-08
9	158	100.0	21	8	US-08-432-Sequence 3, Applicatio	1.00e-08
10	158	100.0	21	5	US-08-161-Sequence 66, Applicati	1.00e-08
11	158	100.0	21	6	US-08-245-Sequence 14, Applicatio	1.00e-08
12	158	100.0	21	15	US-09-089-Sequence 5, Applicatio	1.00e-08
13	158	100.0	22	10	US-08-577-Sequence 2, Applicatio	1.00e-08
14	158	100.0	31	22	US-09-060-Sequence 39, Applicatio	1.00e-08
15	158	100.0	31	22	US-09-060-Sequence 38, Applicatio	1.00e-08
16	158	100.0	31	22	US-09-060-Sequence 42, Applicatio	1.00e-08
17	158	100.0	31	22	US-09-060-Sequence 40, Applicati	1.00e-08
18	158	100.0	31	22	US-09-060-Sequence 41, Applicati	1.00e-08
19	158	100.0	31	5	US-08-161-Sequence 64, Applicati	1.00e-08
20	158	100.0	31	5	US-08-161-Sequence 64, Applicati	1.00e-08
21	158	100.0	32	6	US-08-229-Sequence 14, Applicati	1.00e-08

22	158	100.0	32	1	PCT-US94-0	Sequence 14, Applicati	1.00e-08
23	158	100.0	32	4	US-08-057-	Sequence 14, Applicati	1.00e-08
24	158	100.0	32	9	US-08-488-	Sequence 14, Applicati	1.00e-08
25	158	100.0	158	22	US-09-060-	Sequence 15, Applicati	1.00e-08
26	158	100.0	158	22	US-09-060-	Sequence 14, Applicati	1.00e-08
27	158	100.0	158	22	US-09-060-	Sequence 14, Applicati	1.00e-08
28	158	100.0	158	22	US-09-060-	Sequence 20, Applicati	1.00e-08
29	158	100.0	158	22	US-09-060-	Sequence 18, Applicati	1.00e-08
30	158	100.0	1315	14	US-08-913-	Sequence 1, Applicatio	1.00e-08
31	152	96.2	22	7	US-08-328-	Sequence 8, Applicatio	4.89e-08
32	152	96.2	22	6	US-08-229-	Sequence 5, Applicatio	4.89e-08
33	152	96.2	22	6	US-08-218-	Sequence 8, Applicatio	4.89e-08
34	152	96.2	22	1	PCT-US94-0	Sequence 5, Applicatio	4.89e-08
35	152	96.2	22	9	US-08-488-	Sequence 5, Applicatio	4.89e-08
36	152	96.2	22	14	US-08-926-	Sequence 8, Applicatio	4.89e-08
37	152	96.2	22	4	US-08-060-	Sequence 8, Applicatio	4.89e-08
38	152	96.2	22	4	US-08-057-	Sequence 5, Applicatio	4.89e-08
39	152	96.2	22	12	US-08-718-	Sequence 8, Applicatio	4.89e-08
40	152	96.2	32	7	US-08-328-	Sequence 19, Applicati	4.89e-08
41	152	96.2	32	12	US-08-718-	Sequence 19, Applicati	4.89e-08
42	152	96.2	32	6	US-08-218-	Sequence 19, Applicati	4.89e-08
43	152	96.2	32	4	US-08-060-	Sequence 19, Applicati	4.89e-08
44	152	96.2	34	12	US-08-718-	Sequence 38, Applicati	4.89e-08
45	152	96.2	34	7	US-08-328-	Sequence 38, Applicati	4.89e-08

ALIGNMENTS

RESULT 1
ID US-08-432-483A-3 STANDARD; PRG: 21 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 3, Application US/08432483A
XX
CC Sequence 3, Application US/08432483A
CC GENERAL INFORMATION:
CC APPLICANT: Rittershaus, Charles W.
CC TITLE OF INVENTION: MODULATION OF CHOLESTERYL ESTER
CC TITLE OF INVENTION: TRANSFER PROTEIN (CTP) ACTIVITY
CC NUMBER OF SEQUENCES: 5
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Banner & Witcoff, Ltd.
CC STREET: Ten South Wacker Drive
CC CITY: Chicago
CC STATE: Illinois
CC COUNTRY: USA
CC ZIP: 60606-7407
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: diskette
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Wordperfect 6.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/432,483A
CC FILING DATE: 1-May-1995
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER:
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Leon R. Yankwich
CC REGISTRATION NUMBER: 30,237
CC REFERENCE/DOCKET NUMBER: 95,179(ICS-95179)
CC INFORMATION FOR SEQ ID NO: 3:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 21 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC HYPOTHETICAL:


```

Query Match      100.0%; Score 158; DB 3; Length 21;
Best Local Similarity 100.0%; Pred. No. 1.00e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      1 FNNFTVSFWLRVPKVSASHLE 21
QY      1 FNNFTVSFWLRVPKVSASHLE 21

RESULT 4
ID US-09-049-847-2 STANDARD; PRT; 21 AA.
XX
AC xxxxxx
XX
DT
DE
Sequence 2, Application US/09049847
CC
Sequence 2, Application US/09049847
CC
GENERAL INFORMATION:
CC
APPLICANT: Bay, Sylvie
CC
APPLICANT: Cantacuzene, Daniele
CC
APPLICANT: Leclerc, Claude
CC
APPLICANT: Lo-Map, Richard
CC
TITLE OF INVENTION: Multiple antigen alycopeptide carbohydrate, vaccine
CC
TITLE OF INVENTION: comprising the same and use thereof
CC
FILE REFERENCE: 102.156A
CC
CURRENT APPLICATION NUMBER: US/09/049,847
CC
CURRENT FILING DATE: 1998-03-27
CC
EARLIER APPLICATION NUMBER: 60/041,726
CC
EARLIER FILING DATE: 1997-03-27
CC
NUMBER OF SEQ ID NOS: 4
CC
SOFTWARE: PatentIn Ver. 2.0
CC
SEQ ID NO 2
CC
LENGTH: 21
CC
TYPE: PRT
CC
ORGANISM: Clostridium tetani
CC
SEQUENCE 21 AA; 2479 MW; 2810 CN;

Query Match      100.0%; Score 158; DB 15; Length 21;
Best Local Similarity 100.0%; Pred. No. 1.00e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      1 FNNFTVSFWLRVPKVSASHLE 21
QY      1 FNNFTVSFWLRVPKVSASHLE 21

RESULT 5
ID US-09-171-969-10 STANDARD; PRT; 21 AA.
XX
AC xxxxxx
XX
DT
DE
Sequence 10, Application US/09171969
CC
Sequence 10, Application US/09171969
CC
GENERAL INFORMATION:
CC
APPLICANT: Thomas, Lawrence J.
CC
TITLE OF INVENTION: PLASMID-BASED VACCINE FOR TREATING ATHEROSCLEROSIS
CC
NUMBER OF SEQUENCES: 10
CC
CORRESPONDENCE ADDRESS:
CC
ADDRESSEE: Banner & Witcoff, Ltd.
CC
STREET: 75 State Street, Suite 2300
CC
CITY: Boston
CC
STATE: Massachusetts
CC
COUNTRY: USA
CC
ZIP: 02109-1807
CC
COMPUTER READABLE FORM:
CC
MEDIUM TYPE: Floppy disk
CC
COMPUTER: IBM PC compatible
CC
OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

SOFTWARE: WordPerfect 6.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/171,969
FILING DATE: 01 May 1997 (01.05.97)
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/640,713
FILING DATE: 01 May 1996 (01.05.96)
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/802,967
FILING DATE: 21 February 1997 (21.02.97)
ATTORNEY/AGENT INFORMATION:
NAME: Leon R. Yankwich
REGISTRATION NUMBER: 30,237
REFERENCE/DOCKET NUMBER: TCS 414.1 PCT (05872)
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEetical:
ANTI-SENSE:
FEATURE:
NAME/KEY:
LOCATION:
SEQUENCE 21 AA; 2479 MW; 2810 CN;

Query Match      100.0%; Score 158; DB 16; Length 21;
Best Local Similarity 100.0%; Pred. No. 1.00e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      1 FNNFTVSFWLRVPKVSASHLE 21
QY      1 FNNFTVSFWLRVPKVSASHLE 21

RESULT 6
ID US-08-724-774B-5 STANDARD; PRT; 21 AA.
XX
AC xxxxxx
XX
DT
DE
Sequence 5, Application US/08724774B
CC
Sequence 5, Application US/08724774B
CC
GENERAL INFORMATION:
CC
APPLICANT: Rimoldi, Donata; Jongeneel, Victor; Coulie,
CC
APPLICANT: Pierre; Cerrottini, Jean-Charles; Carrel,
CC
APPLICANT: Stefan; Reed, Daryl
CC
TITLE OF INVENTION: MAGE-10 ENCODING cDNA, The Tumor
CC
TITLE OF INVENTION: Rejection Antigen Precursors Mage-10,
CC
TITLE OF INVENTION: Antibodies Specific To The Molecule, and
CC
NUMBER OF SEQUENCES: 5
CC
CORRESPONDENCE ADDRESS:
CC
ADDRESSEE: Felfe & Lynch
CC
STREET: 805 Third Avenue
CC
CITY: New York City
CC
STATE: New York
CC
COUNTRY: USA
CC
ZIP: 10022
CC
COMPUTER READABLE FORM:
CC
MEDIUM TYPE: Diskette, 3.5 inch, 144 kb storage
CC
COMPUTER: IBM
CC
OPERATING SYSTEM: PC-DOS
CC
SOFTWARE: WordPerfect
CC
CURRENT APPLICATION DATA:
CC
APPLICATION NUMBER: US/08/724,774B
CC
FILING DATE: 03-October-1996
CC
CLASSIFICATION: 530
CC
ATTORNEY/AGENT INFORMATION:

```


CORRESPONDENCE ADDRESS:
ADDRESSES: Pelie & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
COUNTRY: USA
ZIP: 10022

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3
COMPUTER: IBM

TELEPHONE: (202) 672-5399
TELEFAX: (202) 672-5399

CC TELEX: 904136
 CC INFORMATION FOR SEQ ID NO: 2:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 22 amino acids
 CC TYPE: amino acid
 CC STRANDEDNESS: single
 CC TOPOLOGY: linear
 CC SEQUENCE 22 AA: 2582 MW; 2920 CN;

Query Match 100.0%; Score 158; DB 10; Length 22;
 Best Local Similarity 100.0%; Pred. No. 1.00e-08;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 FNNETVSEFWLRVPKVSASHLE 21
 QY 1 FNNETVSEFWLRVPKVSASHLE 21

RESULT 14
 ID US-09-060-294-39 STANDARD; PRT; 31 AA.
 XX
 AC xxxxxx
 XX
 DT
 XX

Sequence 39, Application US/09060294

Sequence 39, Application US/09060294

GENERAL INFORMATION:
 APPLICANT: Martin Roland Jensen
 APPLICANT: Soren Mouritsen
 APPLICANT: Henrik Elsnær
 APPLICANT: Iben Dalum

CC TITLE OF INVENTION: Modified human TNF-alpha molecules, DNA
 CC TITLE OF INVENTION: encoding them, and vaccines containing said modified
 CC NUMBER OF SEQUENCES: 42
 CC CORRESPONDENCE ADDRESS:
 CC ADDRESSEE: Farmaceutisk Laboratorium Ferring A/S
 CC STREET: Indertoften 10
 CC CITY: Vanløse
 CC COUNTRY: Denmark
 CC ZIP: DK-2720

CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Floppy disk
 CC COMPUTER: IBM PC compatible
 CC OPERATING SYSTEM: PC-DOS/MS-DOS
 CC SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: US/09/060,294
 CC FILING DATE: 15-APR-1998
 CC PRIOR APPLICATION DATA:
 CC APPLICATION NUMBER: 60/044,187
 CC FILING DATE: 24-APR-1997
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: Price, D. Douglas
 CC REGISTRATION NUMBER: 24,514
 CC REFERENCE/DOCKET NUMBER: P60953US1
 CC TELEPHONE: (202) 638-6666
 CC TELEFAX: (202) 393-5350
 CC INFORMATION FOR SEQ ID NO: 39:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 31 amino acids
 CC TYPE: amino acid
 CC STRANDEDNESS:
 CC TOPOLOGY: linear

CC MOLECULE TYPE: peptide
 CC HYPOTHETICAL: NO
 CC FRAGMENT TYPE: internal
 CC ORIGINAL SOURCE:
 CC ORGANISM: Homo sapiens
 CC FEATURE:
 CC NAME/KEY: Peptide

CC NAME/KEY: Peptide
 CC LOCATION: 1..31
 CC OTHER INFORMATION: /label= Pep30-2
 CC OTHER INFORMATION: /note= "Pep30-2 is a synthetically prepared truncated
 CC OTHER INFORMATION: form of a TNF-alpha analog comprising the human 1 cell
 CC OTHER INFORMATION: epitope P30 and flanking portions of human TNF-alpha"
 CC SEQUENCE 31 AA: 3577 MW; 6013 CN;

Query Match 100.0%; Score 158; DB 22; Length 31;
 Best Local Similarity 100.0%; Pred. No. 1.00e-08;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 6 FNNETVSEFWLRVPKVSASHLE 26
 QY 1 FNNETVSEFWLRVPKVSASHLE 21

RESULT 15
 ID US-09-060-294-38 STANDARD; PRT; 31 AA.
 XX
 AC xxxxxx
 XX
 DT
 XX

Sequence 38, Application US/09060294

Sequence 38, Application US/09060294

GENERAL INFORMATION:
 APPLICANT: Martin Roland Jensen
 APPLICANT: Soren Mouritsen
 APPLICANT: Henrik Elsnær
 APPLICANT: Iben Dalum

CC TITLE OF INVENTION: Modified human TNF-alpha molecules, DNA
 CC TITLE OF INVENTION: encoding them, and vaccines containing said modified
 CC NUMBER OF SEQUENCES: 42
 CC CORRESPONDENCE ADDRESS:
 CC ADDRESSEE: Farmaceutisk Laboratorium Ferring A/S
 CC STREET: Indertoften 10
 CC CITY: Vanløse
 CC COUNTRY: Denmark
 CC ZIP: DK-2720

CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Floppy disk
 CC COMPUTER: IBM PC compatible
 CC OPERATING SYSTEM: PC-DOS/MS-DOS
 CC SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: US/09/060,294
 CC FILING DATE: 15-APR-1998
 CC PRIOR APPLICATION DATA:
 CC APPLICATION NUMBER: 60/044,187
 CC FILING DATE: 24-APR-1997
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: Price, D. Douglas
 CC REGISTRATION NUMBER: 24,514
 CC REFERENCE/DOCKET NUMBER: P60953US1
 CC TELEPHONE: (202) 638-6666
 CC TELEFAX: (202) 393-5350
 CC INFORMATION FOR SEQ ID NO: 38:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 31 amino acids
 CC TYPE: amino acid
 CC STRANDEDNESS:
 CC TOPOLOGY: linear

CC MOLECULE TYPE: peptide
 CC HYPOTHETICAL: NO
 CC FRAGMENT TYPE: internal
 CC ORIGINAL SOURCE:
 CC ORGANISM: Homo sapiens
 CC FEATURE:
 CC NAME/KEY: Peptide

CC LOCATION: 1.31
CC OTHER INFORMATION: /label= Pep30-1
CC OTHER INFORMATION: /note= Pep30-1 is a synthetically prepared truncated
CC OTHER INFORMATION: form of a TNF-alpha analog comprising human T cell
CC OTHER INFORMATION: epitope p30 and flanking portions of human TNF-alpha"
CC SEQUENCE 31 AA; 3576 MW; 4576 CN;
SQ
Query Match 100.0%; Score 158; DB 22; Length 31;
Best Local Similarity 100.0%; Pred. No. 1.00e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 6 FNNFTVSFWLRVPKVSASHLE 26
QY 1 FNNFTVSFWLRVPKVSASHLE 21
RESULT 16
ID US-09-060-294-42 STANDARD; PRI: 31 AA.
XX xxxxxx
AC
XX
XX
DT
XX
XX
DE Sequence 42, Application US/09060294
XX
Sequence 42, Application US/09060294
CC GENERAL INFORMATION:
CC APPLICANT: Martin Roland Jensen
CC APPLICANT: Soren Mouritsen
CC APPLICANT: Henrik Elsnær
CC APPLICANT: Iben Dalum
CC TITLE OF INVENTION: Modified human TNF-alpha molecules, DNA
CC TITLE OF INVENTION: encoding them, and vaccines containing said modified
CC TITLE OF INVENTION: TNF-alpha or DNA
CC NUMBER OF SEQUENCES: 42
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Farmaceutisk Laboratorium Ferring A/S
CC STREET: Indertoften 10
CC CITY: Vanløse
CC COUNTRY: Denmark
CC ZIP: DK-2720
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/09/060,294
CC FILING DATE: 15-APR-1998
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 60/044,187
CC FILING DATE: 24-APR-1997
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Price, D. Douglas
CC REGISTRATION NUMBER: 24,514
CC REFERENCE/DOCKET NUMBER: P60953US1
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (202) 638-6666
CC TELEFAX: (202) 393-5350
CC INFORMATION FOR SEQ ID NO: 42:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 31 amino acids
CC TYPE: amino acid
CC STRANDEDNESS:
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC HYPOTHETICAL: NO
CC FRAGMENT TYPE: internal
CC ORIGINAL SOURCE:
CC ORGANISM: Homo sapiens
CC FEATURE:
CC NAME/KEY: Peptide
CC LOCATION: 1.31
CC OTHER INFORMATION: /label= Pep30-1

CC OTHER INFORMATION: /label= Pep30-5
CC OTHER INFORMATION: /note= Pep30-5 is a synthetically prepared truncated
CC OTHER INFORMATION: form of a TNF-alpha analog comprising the human T cell
CC OTHER INFORMATION: epitope p30 and flanking portions of human TNF-alpha"
CC SEQUENCE 31 AA; 3532 MW; 5343 CN;
SQ
Query Match 100.0%; Score 158; DB 22; Length 31;
Best Local Similarity 100.0%; Pred. No. 1.00e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 6 FNNFTVSFWLRVPKVSASHLE 26
QY 1 FNNFTVSFWLRVPKVSASHLE 21
RESULT 17
ID US-09-060-294-40 STANDARD; PRT: 31 AA.
XX xxxxxx
AC
XX
XX
DT
XX
XX
DE Sequence 40, Application US/09060294
XX
Sequence 40, Application US/09060294
CC GENERAL INFORMATION:
CC APPLICANT: Martin Roland Jensen
CC APPLICANT: Soren Mouritsen
CC APPLICANT: Henrik Elsnær
CC APPLICANT: Iben Dalum
CC TITLE OF INVENTION: Modified human TNF-alpha molecules, DNA
CC TITLE OF INVENTION: encoding them, and vaccines containing said modified
CC TITLE OF INVENTION: TNF-alpha or DNA
CC NUMBER OF SEQUENCES: 42
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Farmaceutisk Laboratorium Ferring A/S
CC STREET: Indertoften 10
CC CITY: Vanløse
CC COUNTRY: Denmark
CC ZIP: DK-2720
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/09/060,294
CC FILING DATE: 15-APR-1998
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 60/044,187
CC FILING DATE: 24-APR-1997
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Price, D. Douglas
CC REGISTRATION NUMBER: 24,514
CC REFERENCE/DOCKET NUMBER: P60953US1
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (202) 638-6666
CC TELEFAX: (202) 393-5350
CC INFORMATION FOR SEQ ID NO: 40:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 31 amino acids
CC TYPE: amino acid
CC STRANDEDNESS:
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC HYPOTHETICAL: NO
CC FRAGMENT TYPE: internal
CC ORIGINAL SOURCE:
CC ORGANISM: Homo sapiens
CC FEATURE:
CC NAME/KEY: Peptide
CC LOCATION: 1.31
CC OTHER INFORMATION: /label= Pep30-3

ID US-08-161-889A-64 STANDARD; PRT; 31 AA.
XX
AC xxxxxx
DT
DT
XX
DE Sequence 64, Application US/08161889A
XX
CC Sequence 64, Application US/08161889A
CC GENERAL INFORMATION:
CC APPLICANT: Geysen, H. Mario
CC APPLICANT: Rodda, Stuart J.
CC TITLE OF INVENTION: T-Cell Epitopes
CC NUMBER OF SEQUENCES: 74
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Grant D. Green
CC STREET: 4560 Horton St.
CC CITY: Emeryville
CC STATE: CA
CC COUNTRY: USA
CC ZIP: 94608
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC OPERATING SYSTEM: IBM PC compatible
CC SOFTWARE: PatentIn Release #1.0, Version #1.30B
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/161,889A
CC FILING DATE: 02-DEC-1993
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/984,852
CC FILING DATE: 02-DEC-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Green, Grant D.
CC REGISTRATION NUMBER: 31,259
CC REFERENCE/DOCKET NUMBER: 0222.002
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 510-601-2706
CC TELEFAX: 510-655-3542
CC INFORMATION FOR SEQ ID NO: 64:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 31 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: Peptide
CC SEQUENCE 31 AA; 3694 MW; 5041 CN;
SQ
Query Match 100.0%; Score 158; DB 5; Length 31;
Best Local Similarity 100.0%; Pred. No. 1.00e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 7 FNNFTVSFWLRVPKVSASHLE 27
QY 1 FNNFTVSFWLRVPKVSASHLE 21
RESULT 21
ID US-08-229-275-14 STANDARD; PRT; 32 AA.
XX
AC xxxxxx
XX
XX
DT
DT
XX
DE Sequence 14, Application US/08229275
XX
CC Sequence 14, Application US/08229275
CC GENERAL INFORMATION:
CC APPLICANT: Ladd, Anna
CC APPLICANT: Wang, Chang Yi
CC APPLICANT: Zamb, Timothy
CC TITLE OF INVENTION: Immunogenic LHRH peptide constructs as

CC TITLE OF INVENTION: vaccines for treatment of prostate cancer and induction
CC TITLE OF INVENTION: infertility
CC NUMBER OF SEQUENCES: 53
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: M. Lisa Wilson
CC STREET: 25 Davids Drive
CC CITY: Hauppauge
CC STATE: NY
CC COUNTRY: US
CC ZIP: 11788
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC OPERATING SYSTEM: IBM PC compatible
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/229,275
CC FILING DATE: 13-APR-1994
CC CLASSIFICATION: 424
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Wilson, M L
CC REGISTRATION NUMBER: 34,045
CC REFERENCE/DOCKET NUMBER: 2003Z
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (516)273-2828
CC TELEFAX: (516)273-1717
CC INFORMATION FOR SEQ ID NO: 14:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 32 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 32 AA; 3789 MW; 6283 CN;
SQ
Query Match 100.0%; Score 158; DB 6; Length 32;
Best Local Similarity 100.0%; Pred. No. 1.00e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 3 FNNFTVSFWLRVPKVSASHLE 23
QY 1 FNNFTVSFWLRVPKVSASHLE 21
RESULT 22
ID PCT-US94-04832A-14 STANDARD; PRT; 32 AA.
XX
AC xxxxxx
XX
DT
DT
XX
DE Sequence 14, Application PC/TUS9404832A
XX
CC Sequence 14, Application PC/TUS9404832A
CC GENERAL INFORMATION:
CC APPLICANT: Ladd, Anna
CC APPLICANT: Wang, Chang Yi
CC APPLICANT: Zamb, Timothy
CC TITLE OF INVENTION: Immunogenic LHRH peptide constructs
CC TITLE OF INVENTION: and synthetic universal immune stimulators for vaccines
CC NUMBER OF SEQUENCES: 114
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE:
CC STREET: 400 Garden City Plaza
CC CITY: Garden City
CC STATE: NY
CC COUNTRY: US
CC ZIP: 11530
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC OPERATING SYSTEM: IBM PC compatible
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC

CC APPLICATION NUMBER: PCT/US94/04832A
CC FILING DATE: 13-APR-1994
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME:
CC REGISTRATION NUMBER:
CC REFERENCE/DOCKET NUMBER:
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (516)742-4343
CC TELEFAX: (516)742-4366
CC INFORMATION FOR SEQ ID NO: 14:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 32 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 32 AA: 3789 MW: 6283 CN;
SQ
Query Match 100.0%; Score 158; DB 1; Length 32;
Best Local Similarity 100.0%; Pred. No. 1.00e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 3 FNNFTVSFWLRVPKVSASHLE 23
QY 1 FNNFTVSFWLRVPKVSASHLE 21
RESULT 23
ID US-08-057-156-14 STANDARD; PRT: 32 AA.
XX
AC xxxxxx
XX
DT
XX
DE
XX
Sequence 14, Application US/08057166
Sequence 14, Application US/08057166
GENERAL INFORMATION:
APPLICANT: Ladd, Anna
APPLICANT: Wang, Chang Yi
TITLE OF INVENTION: Immunogenic Peptides Which Contain LHRH
TITLE OF INVENTION: and a Helper T Cell Epitope for Treatment of Prostate
TITLE OF INVENTION: Cancer and Induction of Infertility
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: M. Lisa Wilson, United Biomedical Inc.
STREET: 25 Davids Dr.
CITY: Hauppauge
STATE: NY
COUNTRY: U.S.A.
ZIP: 11788
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/057,166
FILING DATE: 19930427
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Wilson, M. Lisa
REGISTRATION NUMBER: 34,045
REFERENCE/DOCKET NUMBER: 2003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516)273-2828
TELEFAX: (516)273-1717
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 32 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: peptide

SQ SEQUENCE 32 AA: 3789 MW: 6283 CN;
Query Match 100.0%; Score 158; DB 4; Length 32;
Best Local Similarity 100.0%; Pred. No. 1.00e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 3 FNNFTVSFWLRVPKVSASHLE 23
QY 1 FNNFTVSFWLRVPKVSASHLE 21
RESULT 24
ID US-08-488-320A-14 STANDARD; PRT: 32 AA.
XX
AC xxxxxx
XX
DT
XX
DE
XX
Sequence 14, Application US/08488320A
Sequence 14, Application US/08488320A
GENERAL INFORMATION:
APPLICANT: Ladd, Anna
APPLICANT: Wang, Chang Yi
TITLE OF INVENTION: Immunogenic Peptides Which Contain LHRH
TITLE OF INVENTION: And A Helper T-Cell Epitope For Treatment Of Prostate C
TITLE OF INVENTION: And Induction Of Infertility
NUMBER OF SEQUENCES: 114
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: U.S.A.
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/488,320A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/488,351
FILING DATE: 07-JUN-1995
APPLICATION NUMBER: 08/446,692
FILING DATE: 07-JUN-1995
APPLICATION NUMBER: 08/229,275
FILING DATE: 14-APR-1994
APPLICATION NUMBER: 08/057,166
FILING DATE: 27-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Maria C. H. Lin
REGISTRATION NUMBER: 29,323
REFERENCE/DOCKET NUMBER: 1151-4146USA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)758-4800
TELEFAX: (212)751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 32 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE 32 AA: 3789 MW: 6283 CN;
SQ
Query Match 100.0%; Score 158; DB 9; Length 32;
Best Local Similarity 100.0%; Pred. No. 1.00e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Db 3 FNNFTVSEWLRVPKVSASHLE 23
   |||||||
QY 1 FNNFTVSEWLRVPKVSASHLE 21

RESULT 25
ID US-09-060-294-16 STANDARD; PRT; 158 AA.
AC xxxxxx
XX
XX
DT
XX
DE
DE
XX
XX
XX
XX
Sequence 16, Application US/09060294
Sequence 16, Application US/09060294
GENERAL INFORMATION:
APPLICANT: Martin Roland Jensen
APPLICANT: Soren Mouritsen
APPLICANT: Henrik Elsnær
APPLICANT: Iben Dalum
TITLE OF INVENTION: Modified human TNF-alpha molecules, DNA
TITLE OF INVENTION: encoding them, and vaccines containing said modified
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pharmaceutisk Laboratorium Ferring A/S
STREET: Indertoften 10
CITY: Vanløse
COUNTRY: Denmark
ZIP: DK-2720
COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/060,294
FILING DATE: 15-APR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/044,187
FILING DATE: 24-APR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Price, D. Douglas
REGISTRATION NUMBER: 24,514
REFERENCE/DOCKET NUMBER: P60953US1
TELEPHONE: (202) 638-6666
TELEFAX: (202) 393-5350
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 158 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE 158 AA; 17696 MW; 131187 CN;

Query Match 100.0%; Score 158; DB 22; Length 158;
Best Local Similarity 100.0%; Pred. No. 1.00e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 65 FNNFTVSEWLRVPKVSASHLE 85
   |||||||
QY 1 FNNFTVSEWLRVPKVSASHLE 21

RESULT 26
ID US-09-060-294-12 STANDARD; PRT; 158 AA.
AC xxxxxx
XX
XX
XX
DT
XX
XX
```

```
DE
XX
CC
CC
Sequence 12, Application US/09060294
Sequence 12, Application US/09060294
GENERAL INFORMATION:
APPLICANT: Martin Roland Jensen
APPLICANT: Soren Mouritsen
APPLICANT: Henrik Elsnær
APPLICANT: Iben Dalum
TITLE OF INVENTION: Modified human TNF-alpha molecules, DNA
TITLE OF INVENTION: encoding them, and vaccines containing said modified
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pharmaceutisk Laboratorium Ferring A/S
STREET: Indertoften 10
CITY: Vanløse
COUNTRY: Denmark
ZIP: DK-2720
COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/060,294
FILING DATE: 15-APR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/044,187
FILING DATE: 24-APR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Price, D. Douglas
REGISTRATION NUMBER: 24,514
REFERENCE/DOCKET NUMBER: P60953US1
TELEPHONE: (202) 638-6666
TELEFAX: (202) 393-5350
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 158 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE 158 AA; 17648 MW; 130661 CN;

Query Match 100.0%; Score 158; DB 22; Length 158;
Best Local Similarity 100.0%; Pred. No. 1.00e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 11 FNNFTVSEWLRVPKVSASHLE 31
   |||||||
QY 1 FNNFTVSEWLRVPKVSASHLE 21

RESULT 27
ID US-09-060-294-14 STANDARD; PRT; 158 AA.
AC xxxxxx
XX
XX
DT
XX
DE
DE
XX
XX
Sequence 14, Application US/09060294
Sequence 14, Application US/09060294
GENERAL INFORMATION:
APPLICANT: Martin Roland Jensen
APPLICANT: Soren Mouritsen
APPLICANT: Henrik Elsnær
APPLICANT: Iben Dalum
TITLE OF INVENTION: Modified human TNF-alpha molecules, DNA
TITLE OF INVENTION: encoding them, and vaccines containing said modified
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
```



```
CC      TYPE: amino acid
CC      TOPOLOGY: linear
CC      MOLECULE TYPE: protein
SQ      SEQUENCE 158 AA; 17566 MW; 133193 CN;

Query Match      100.0%; Score 158; DB 22; Length 158;
Best Local Similarity 100.0%; Pred. No. 1.00e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 108 FNNFTVSFWLRVPKVSASHLE 128
QY 1 FNNFTVSFWLRVPKVSASHLE 21
|||||

RESULT 30
ID US-08-913-880-1 STANDARD; PRT; 1315 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 1, Application US/08913880
XX
CC Sequence 1, Application US/08913880
CC GENERAL INFORMATION:
CC APPLICANT: MATSUDA, Morihiro
CC TITLE OF INVENTION: TETANUS TOXIN FUNCTIONAL FRAGMENT ANTIGEN AND TETANUS
CC TITLE OF INVENTION: VACCINE
CC FILE REFERENCE: 216-380P
CC CURRENT APPLICATION NUMBER: US/08/913,880
CC CURRENT FILING DATE: 1997-09-24
CC NUMBER OF SEQ ID NOS: 9
CC SOFTWARE: PatentIn Ver. 2.0
CC SEQ ID NO 1
CC LENGTH: 1315
CC TYPE: PRT
CC ORGANISM: Clostridium tetani
SQ SEQUENCE 1315 AA; 150681 MW; 9110516 CN;

Query Match      100.0%; Score 158; DB 14; Length 1315;
Best Local Similarity 100.0%; Pred. No. 1.00e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 947 FNNFTVSFWLRVPKVSASHLE 967
QY 1 FNNFTVSFWLRVPKVSASHLE 21
|||||
```

Search completed: Tue Aug 17 16:19:03 1999
Job time : 56 secs.

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Tue Aug 17 16:19:20 1999; MasPar time 6.91 Seconds
Tabular output not generated.
Title: >US-09-049-847-2
Description: (1-21) from US09049847.pap
Perfect Score: 158
Sequence: 1 FNNFTVFWLRVPKVSASHLE 21
Scoring table: PAM 150
Gap 15
Searched: 411786 seqs, 50406085 residues
Post-processing: Minimum Match 100%
Listing first 1000 summaries
Maximum DB seq length 21
Database: a-pending
1:PG 2:U60 3:J7 4:U80 5:U81 6:U82 7:U83 8:U84A 9:U84B
10:U85 11:U86 12:U87 13:U88 14:U89 15:U90 16:U91 17:U92
18:U93 19:NEWP 20:NEWU6 21:NEWU8 22:NEWU9
Statistics: Mean 22.277; Variance 74.265; scale 0.300
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
SUMMARIES
Result No. Score Query Match Length DB ID Description Pred. No.
1 158 100.0 21 8 US-08-432- Sequence 3, Applicatio 1.00e-08
2 158 100.0 21 14 US-08-945- Sequence 3, Applicatio 1.00e-08
3 158 100.0 21 3 US-07-678- Sequence 1, Applicatio 1.00e-08
4 158 100.0 21 15 US-09-049- Sequence 2, Applicatio 1.00e-08
5 158 100.0 21 16 US-09-171- Sequence 10, Applicati 1.00e-08
6 158 100.0 21 12 US-08-724- Sequence 5, Applicatio 1.00e-08
7 158 100.0 21 8 US-08-406- Sequence 4, Applicatio 1.00e-08
8 158 100.0 21 5 US-08-161- Sequence 66, Applicati 1.00e-08
9 158 100.0 21 8 US-08-432- Sequence 3, Applicatio 1.00e-08
10 158 100.0 21 5 US-08-161- Sequence 66, Applicati 1.00e-08
11 158 100.0 21 6 US-08-245- Sequence 14, Applicati 1.00e-08
12 158 100.0 21 15 US-09-089- Sequence 5, Applicatio 1.00e-08
Note: Post-processor removed 988 summaries from list due to search parameters chosen.
ALIGNMENTS
RESULT 1
ID US-08-432-483A-3 STANDARD; PRT; 21 AA.

xx
AC
DT
XX
DE
XX
Sequence 3, Application US/08432483A
Sequence 3, Application US/08432483A
GENERAL INFORMATION:
APPLICANT: Rittershaus, Charles W.
TITLE OF INVENTION: MODULATION OF CHOLESTERYL ESTER
TITLE OF INVENTION: TRANSFER PROTEIN (CETP) ACTIVITY
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Witcoff, Ltd.
STREET: Ten South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606-7407
COMPUTER READABLE FORM:
MEDIUM TYPE: diskette
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WordPerfect 6.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/432,483A
FILING DATE: 1-May-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Leon R. Yankwich
REGISTRATION NUMBER: 30,237
REFERENCE/DOCKET NUMBER: 95,179(TCS-95179)
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL:
ANTI-SENSE:
FEATURE:
NAME/KEY: 21-amino acid tetanus toxoid universal
NAME/KEY: helper T cell epitope.
LOCATION:
PUBLICATION INFORMATION:
AUTHORS: Panina-Bordignon, P., et al.
TITLE: Universally immunogenic T cell
TITLE: epitopes: promiscuous binding to human MHC class II and
JOURNAL: European Journal of Immunology
VOLUME: 19
ISSUE:
PAGES: 2237-2242
DATE: 1989
RELEVANT RESIDUES IN SEQ ID NO: 3: FROM 1 TO 21
SEQUENCE 21 AA; 2479 MW; 2810 CN;
Query Match 100.0%; Score 158; DB 8; Length 21;
Best Local Similarity 100.0%; Pred.No. 1.00e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 1 FNNFTVFWLRVPKVSASHLE 21
QY 1 FNNFTVFWLRVPKVSASHLE 21
RESULT 2
ID US-08-945-289-3 STANDARD; PRT; 21 AA.
XX

```
AC          xxxxxx
XX
DT
DE
XX
XX
XX
Sequence 3, Application US/08945289
CC
CC
CC
Sequence 3, Application US/08945289
CC
CC
CC
GENERAL INFORMATION:
CC
CC
CC
APPLICANT: Rittershaus, Charles, W.
CC
CC
CC
APPLICANT: Thomas, Lawrence J.
CC
CC
CC
TITLE OF INVENTION: MODULATION OF CHOLESTERYL ESTER
CC
CC
CC
TITLE OF INVENTION: TRANSFER PROTEIN (CTP) ACTIVITY
CC
CC
CC
NUMBER OF SEQUENCES: 9
CC
CC
CC
CORRESPONDENCE ADDRESS:
CC
CC
CC
ADDRESSEE: Yankwich & Associates
CC
CC
CC
STREET: 130 Bishop Allen Drive
CC
CC
CC
CITY: Cambridge
CC
CC
CC
STATE: Massachusetts
CC
CC
CC
COUNTRY: USA
CC
CC
CC
ZIP: 02139
CC
CC
CC
COMPUTER READABLE FORM:
CC
CC
CC
MEDIUM TYPE: Floppy disk
CC
CC
CC
COMPUTER: IBM PC compatible
CC
CC
CC
OPERATING SYSTEM: Windows 95
CC
CC
CC
SOFTWARE: Word 97
CC
CC
CC
CURRENT APPLICATION DATA:
CC
CC
CC
APPLICATION NUMBER: US/08/945,289
CC
CC
CC
FILING DATE: October 17, 1997
CC
CC
CC
CLASSIFICATION: 514
CC
CC
CC
PRIOR APPLICATION DATA:
CC
CC
CC
APPLICATION NUMBER: 08/432,483
CC
CC
CC
FILING DATE: May 1, 1995
CC
CC
CC
ATTORNEY/AGENT INFORMATION:
CC
CC
CC
NAME: Leon R. Yankwich
CC
CC
CC
REGISTRATION NUMBER: 30,237
CC
CC
CC
REFERENCE/DOCKET NUMBER: TCS-411.1P US
CC
CC
CC
INFORMATION FOR SEQ ID NO: 3:
CC
CC
CC
SEQUENCE CHARACTERISTICS:
CC
CC
CC
LENGTH: 21 amino acids
CC
CC
CC
TYPE: amino acid
CC
CC
CC
TOPOLOGY: linear
CC
CC
CC
MOLECULE TYPE: peptide
CC
CC
CC
HYPOTHETICAL:
CC
CC
CC
ANTI-SENSE:
CC
CC
CC
FEATURE:
CC
CC
CC
NAME/KEY: 21-amino acid tetanus toxoid universal helper T cell epitope.
CC
CC
CC
LOCATION:
CC
CC
CC
PUBLICATION INFORMATION:
CC
CC
CC
AUTHORS: Paina-Bordignon, P., et al.
CC
CC
CC
TITLE: Universally immunogenic T cell epitopes: promiscuous binding to human
CC
CC
CC
JOURNAL: European Journal of Immunology
CC
CC
CC
VOLUME: 19
CC
CC
CC
ISSUE:
CC
CC
CC
PAGES: 2237-2242
CC
CC
CC
DATE: 1989
CC
CC
CC
RELEVANT RESIDUES IN SEQ ID NO: 3: FROM 1 TO 21
CC
CC
CC
SEQUENCE 21 AA; 2479 MW; 2810 CN;

Query Match 100.0%; Score 158; DB 14; Length 21;
Best Local Similarity 100.0%; Pred. No. 1.00e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 FNNFTVSFWLRPKVSASHLE 21
QY 1 FNNFTVSFWLRPKVSASHLE 21

RESULT 3
ID US-07-678-684B-1 STANDARD; PRT; 21 AA.
XX
XX
AC xxxxxx
XX
XX
DT
DE
XX
XX
Sequence 2, Application US/09049847
CC
CC
CC
GENERAL INFORMATION:
CC
CC
CC
APPLICANT: Bay, Sylvie
CC
CC
CC
APPLICANT: Cantacuzene, Daniele
CC
CC
CC
APPLICANT: Leclerc, Claude
CC
CC
CC
APPLICANT: Lo-Man, Richard
CC
CC
CC
TITLE OF INVENTION: Multiple antigen alycoepitide carbohydrate, vaccine
CC
CC
CC
TITLE OF INVENTION: comprising the same and use thereof
CC
CC
CC
FILE REFERENCE: 102.166A
CC
CC
CC
CURRENT APPLICATION NUMBER: US/09/049,847
CC
CC
CC
CURRENT FILING DATE: 1998-03-27
CC
CC
CC
EARLIER APPLICATION NUMBER: 60/041,726
```

CC EARLIER FILING DATE: 1997-03-27
 CC NUMBER OF SEQ ID NOS: 4
 CC SOFTWARE: Patentin Ver. 2.0
 CC SEQ ID NO 2
 CC LENGTH: 21
 CC TYPE: PRT
 CC ORGANISM: Clostridium tetani
 CC SEQ SEQUENCE 21 AA; 2479 MW; 2810 CN;

Query Match 100.0%; Score 158; DB 15; Length 21;
 Best Local Similarity 100.0%; Pred. No. 1.00e-08;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 FNNFTVSFWLRVPKVSASHLE 21
 |||||
 QY 1 FNNFTVSFWLRVPKVSASHLE 21

RESULT 5
 ID US-09-171-969-10 STANDARD; PRT; 21 AA.
 XX
 AC xxxxxx
 XX
 XX
 DT
 XX

Sequence 10, Application US/09171969

Sequence 10, Application US/09171969

GENERAL INFORMATION:
 APPLICANT: Thomas, Lawrence J.
 TITLE OF INVENTION: PLASMID-BASED VACCINE FOR TREATING ATHEROSCLEROSIS
 NUMBER OF SEQUENCES: 10
 CORRESPONDENCE ADDRESS:

ADDRESSEE: Banner & Witcoff, Ltd.
 STREET: 75 State Street, Suite 2300
 CITY: Boston

STATE: Massachusetts

COUNTRY: USA

ZIP: 02109-1807

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WordPerfect 6.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/171,969

FILING DATE: 01 May 1997 (01.05.97)

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/640,713

FILING DATE: 01 May 1996 (01.05.96)

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/802,967

FILING DATE: 21 February 1997 (21.02.97)

ATTORNEY/AGENT INFORMATION:

NAME: Leon R. Yankwich

REGISTRATION NUMBER: 30,237

REFERENCE/DOCKET NUMBER: TCS 414.1 PCT (05872)

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 21 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL:

ANTI-SENSE:

FEATURE:

NAME/KEY:

LOCATION:

SEQ SEQUENCE 21 AA; 2479 MW; 2810 CN;

Query Match 100.0%; Score 158; DB 16; Length 21;
 Best Local Similarity 100.0%; Pred. No. 1.00e-08;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 1 FNNFTVSFWLRVPKVSASHLE 21
 |||||
 QY 1 FNNFTVSFWLRVPKVSASHLE 21

RESULT 6
 ID US-08-724-774B-5 STANDARD; PRT; 21 AA.
 XX
 AC xxxxxx
 XX
 DT
 XX
 XX
 DE
 XX

Sequence 5, Application US/08724774B

Sequence 5, Application US/08724774B

GENERAL INFORMATION:

APPLICANT: Rimoldi, Donata; Jongeneel, Victor; Coulie,

APPLICANT: Pierre; Cerrottini, Jean-Charles; Carrel,

APPLICANT: Stefan; Reed, Daryl

TITLE OF INVENTION: MAGE-10 ENCODING cDNA, The Tumor

TITLE OF INVENTION: Rejection Antigen Precursors Mage-10,

TITLE OF INVENTION: Antibodies Specific To The Molecule, and

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: Felfe & Lynch

STREET: 805 Third Avenue

CITY: New York City

STATE: New York

COUNTRY: USA

ZIP: 10022

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inch, 144 kb storage

COMPUTER: IBM

OPERATING SYSTEM: PC-DOS

SOFTWARE: Wordperfect

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/724,774B

FILING DATE: 03-October-1996

CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:

NAME: Hanson, Norman D.

REGISTRATION NUMBER: 30,946

REFERENCE/DOCKET NUMBER: LUD 5457

TELEPHONE: (212) 688-9200

TELEFAX: (212) 838-3884

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 21

TYPE: amino acid

TOPOLOGY: linear

SEQ SEQUENCE 21 AA; 2479 MW; 2810 CN;

Query Match 100.0%; Score 158; DB 12; Length 21;
 Best Local Similarity 100.0%; Pred. No. 1.00e-08;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 FNNFTVSFWLRVPKVSASHLE 21
 |||||
 QY 1 FNNFTVSFWLRVPKVSASHLE 21

RESULT 7
 ID US-08-406-916B-4 STANDARD; PRT; 21 AA.
 XX
 AC xxxxxx
 XX
 DT
 XX

Sequence 4, Application US/08406916B

NAME: Leon R. Yankwich
REGISTRATION NUMBER: 30,237
REFERENCE/DOCKET NUMBER: 95,179(TCS-95179)
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TYPE: amino acids
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL:
ANTI-SENSE:
FEATURE:
NAME/KEY: 21-amino acid tetanus toxoid
LOCATION:
PUBLICATION INFORMATION:
AUTHORS: Panina-Bordignon, P., et al.
TITLE: Universally immunogenic T cell
TITLE: epitopes: promiscuous binding to human MHC class II and
TITLE: promiscuous recognition by T cells
JOURNAL: European Journal of Immunology
VOLUME: 19
ISSUE:
PAGES: 2237-2242
DATE: 1989
RELEVANT RESIDUES IN SEQ ID NO: 3: FROM 1 TO 21
SEQUENCE 21 AA; 2479 MW; 2810 CN;

Query Match 100.0%; Score 158; DB 8; Length 21;
Best Local Similarity 100.0%; Pred. No. 1,00e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 FNNITVSFWLRVPKVSASHLE 21
1 FNNITVSFWLRVPKVSASHLE 21

QY

RESULT 10
ID US-08-161-889-66 STANDARD; PRT; 21 AA.
XX
AC xxxxxx
AC
DT
DT
XX
XX
DE Sequence 66, Application US/08161889
XX
CC Sequence 66, Application US/08161889
CC GENERAL INFORMATION:
CC APPLICANT: Geysen, H. Mario
CC APPLICANT: Rodda, Stuart J.
CC TITLE OF INVENTION: T-Cell Epitopes
CC NUMBER OF SEQUENCES: 75
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Grant D. Green
CC STREET: 4560 Horton St.
CC CITY: Emeryville
CC STATE: CA
CC COUNTRY: USA
CC ZIP: 94608
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30B
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/161,889
CC FILING DATE: 02-DEC-1993
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/984,852
CC FILING DATE: 02-DEC-1992
CC ATTORNEY/AGENCY INFORMATION:
CC NAME: Green, Grant D.
CC REGISTRATION NUMBER: 31,259

CC ANTI-SENSE: NO
CC ORIGINAL SOURCE:
CC INDIVIDUAL ISOLATE: T antigen, TT3 peptide
SQ SEQUENCE 21 AA; 2479 MW; 2810 CN;
Search completed: Tue Aug 17 16:21:06 1999
Job time : 106 secs.

Query Match 100.0%; Score 158; DB 6; Length 21;
Best Local Similarity 100.0%; Pred. No. 1.00e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 1 FNNFTVSFWLRVPKVSASHLE 21
QY 1 FNNFTVSFWLRVPKVSASHLE 21

RESULT 12
ID US-09-089-595-5 STANDARD; PRT; 21 AA.

XX xxxxxx

Sequence 5, Application US/09089595

Sequence 5, Application US/09089595
GENERAL INFORMATION:
APPLICANT: Rimoldi, Donata; Jongeneel, Victor; Coulie,
APPLICANT: Pierre; Cerrottini, Jean-Charles; Carrel,
APPLICANT: Stefan; Reed, Daryl
TITLE OF INVENTION: MAG-10 ENCODING cDNA, The Tumor
TITLE OF INVENTION: Rejection Antigen Precuros Mage-10,
TITLE OF INVENTION: Antibodies Specific To The Molecule, and
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felte & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
COUNTRY: USA
ZIP: 10022

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 144 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/089,595
FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/724,774
FILING DATE: 03-October-1996
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, Norman D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: IUD 5457
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 21
TYPE: amino acid
TOPOLOGY: linear
SQ SEQUENCE 21 AA; 2479 MW; 2810 CN;

Query Match 100.0%; Score 158; DB 15; Length 21;
Best Local Similarity 100.0%; Pred. No. 1.00e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 FNNFTVSFWLRVPKVSASHLE 21
QY 1 FNNFTVSFWLRVPKVSASHLE 21

M P S R L H
***** (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Tue Aug 17 16:11:13 1999; MasPar time 1.55 Seconds
Tabular output not generated. 137.199 Million cell updates/sec

Title: >US-09-049-847-2
Description: (1-21) from US09049847.pep
Perfect Score: 158
Sequence: 1 FNNFTVFWLRVPKVSASHLE 21

Scoring table: PAM 150
Gap 15

Searched: 106580 seqs, 10152877 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:PCT9_COMB 4:backfiles1

Statistics: Mean 20.415; Variance 71.502; scale 0.286

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	158	100.0	21	2	US-08-450-Sequence 8, Applicatio	3.25e-09
2	158	100.0	21	1	US-07-610-Sequence 1, Applicatio	3.25e-09
3	158	100.0	21	2	US-08-661-Sequence 12, Applicati	3.25e-09
4	158	100.0	21	3	PCT-US93-1Sequence 66, Applicati	3.25e-09
5	158	100.0	31	3	PCT-US93-1Sequence 64, Applicati	3.25e-09
6	158	100.0	32	2	US-08-488-Sequence 14, Applicati	3.25e-09
7	158	100.0	32	2	US-08-446-Sequence 4, Applicatio	3.25e-09
8	158	100.0	452	1	US-07-618-Sequence 2, Applicatio	3.25e-09
9	158	100.0	452	1	US-08-280-Sequence 4, Applicatio	3.25e-09
10	158	100.0	452	1	US-07-618-Sequence 2, Applicatio	3.25e-09
11	158	100.0	452	1	US-08-280-Sequence 8, Applicatio	3.25e-09
12	158	100.0	452	1	US-08-110-Sequence 5, Applicatio	3.25e-09
13	158	100.0	618	2	PCT-US95-1Sequence 8, Applicatio	1.50e-08
14	152	96.2	22	3	PCT-US95-1Sequence 5, Applicatio	1.50e-08
15	152	96.2	22	2	US-08-488-Sequence 5, Applicatio	1.50e-08
16	152	96.2	22	2	US-08-446-Sequence 5, Applicatio	1.50e-08
17	142	89.9	19	1	US-07-610-Sequence 2, Applicatio	1.90e-07
18	133	85.4	20	1	US-08-319-Sequence 11, Applicati	1.10e-06
19	126	79.7	17	1	US-07-610-Sequence 3, Applicatio	1.04e-05
20	116	73.4	14	1	US-08-787-Sequence 43, Applicatio	1.23e-04
21	116	73.4	14	1	US-07-610-Sequence 6, Applicatio	1.23e-04
22	114	72.2	15	1	US-07-610-Sequence 4, Applicatio	2.01e-04
23	114	72.2	15	2	US-08-787-Sequence 44, Applicati	2.01e-04

ALIGNMENTS

RESULT	ID	US-08-450-502-8	STANDARD;	PRT;	21 AA.
XX	AC	xxxxxx			
XX	DT				
XX	XX				
DE	XX	Sequence 8, Application US/08460502			
CC	CC	Sequence 8, Application US/08460502			
CC	CC	Patent No. 5843464			
CC	CC	GENERAL INFORMATION:			
CC	CC	APPLICANT: Bakaletz, Lauren O.			
CC	CC	TITLE OF INVENTION: Synthetic Chimeric Fimbrin Peptides			
CC	CC	NUMBER OF SEQUENCES: 11			
CC	CC	CORRESPONDENCE ADDRESS:			
CC	CC	ADDRESSEE: Calfee, Halter and Griswold			
CC	CC	STREET: 800 Superior Avenue			
CC	CC	CITY: Cleveland			
CC	CC	STATE: Ohio			
CC	CC	COUNTRY: U.S.A.			
CC	CC	ZIP: 44114-2688			
CC	CC	COMPUTER READABLE FORM:			
CC	CC	MEDIUM TYPE: Floppy disk			
CC	CC	COMPUTER: IBM PC compatible			
CC	CC	OPERATING SYSTEM: PC-DOS/MS-DOS			
CC	CC	SOFTWARE: Patent in Release #1.0, Version #1.25			
CC	CC	CURRENT APPLICATION DATA:			
CC	CC	APPLICATION NUMBER: US/08/460,502			
CC	CC	FILING DATE:			
CC	CC	CLASSIFICATION: 424			
CC	CC	ATTORNEY/AGENT INFORMATION:			
CC	CC	NAME: Golrick, Mary E.			
CC	CC	REGISTRATION NUMBER: 34, 829			
CC	CC	REFERENCE/DOCKET NUMBER: 22727/00120			
CC	CC	TELECOMMUNICATION INFORMATION:			
CC	CC	TELEPHONE: (216) 622-8458			
CC	CC	TELEFAX: (216) 241-0816			
CC	CC	INFORMATION FOR SEQ ID NO: 8:			
CC	CC	SEQUENCE CHARACTERISTICS:			
CC	CC	LENGTH: 21 amino acids			
CC	CC	TYPE: amino acid			
CC	CC	STRANDEDNESS: single			
CC	CC	TOPOLOGY: linear			
CC	CC	MOLECULE TYPE: peptide			
SQ	SQ	SEQUENCE 21 AA: 2479 MW; 2810 CN;			

Query Match 100.0%; Score 158; DB 2; Length 21;
Best Local Similarity 100.0%; Pred. No. 3.25e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 FNNFTVSWLRVPKVSASHLE 21
QY 1 FNNFTVSWLRVPKVSASHLE 21

RESULT 2
ID US-07-610-525-1 STANDARD; PRT; 21 AA.

XX xxxxxx

DE Sequence 1, Application US/07610525

XX Patent No. 5196512

CC GENERAL INFORMATION:

CC APPLICANT: BIANCHI Elisabetta

CC APPLICANT: PESSI Antonello

CC APPLICANT: CORRADIN Giampietro

CC TITLE OF INVENTION: SYNTHETIC PEPTIDES USEFUL AS

CC TITLE OF INVENTION: UNIVERSAL CARRIERS FOR THE PREPARATION OF IMMUNOGENIC COMPOUNDS

CC TITLE OF INVENTION: AND THEIR USE IN THE DEVELOPMENT OF SYNTHETIC VACCINES.

CC NUMBER OF SEQUENCES: 7

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: SHEA & GOULD

CC STREET: 1251 AVENUE OF THE AMERICAS

CC CITY: NEW YORK

CC STATE: NEW YORK

CC COUNTRY: UNITED STATES

CC ZIP: 10020-1193

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: Patent in Release #1.24

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/07/610.525

CC FILING DATE: 19901108

CC CLASSIFICATION: 424

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: 212-827-3000

CC TELEFAX: 212-840-6702

CC TELEX: 423973

CC INFORMATION FOR SEQ ID NO: 1:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 21 amino acid residues

CC TYPE: AMINO ACID

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: peptide

CC HYPOTHETICAL: no

CC FRAGMENT TYPE: internal fragment

CC SEQUENCE 21 AA; 2479 MW; 2810 CN;

Query Match 100.0%; Score 158; DB 1; Length 21;
Best Local Similarity 100.0%; Pred. No. 3.25e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 FNNFTVSWLRVPKVSASHLE 21

QY 1 FNNFTVSWLRVPKVSASHLE 21

RESULT 3
ID US-08-661-052-12 STANDARD; PRT; 21 AA.

XX xxxxxx

XX Sequence 12, Application US/08661052
DT
XX
DE
XX
CC Sequence 12, Application US/08661052
CC Patent No. 5837243
CC GENERAL INFORMATION:
CC APPLICANT: Yashwant M. Deo
CC APPLICANT: Joel Goldstein
CC APPLICANT: Robert Graziano
CC APPLICANT: Chezian Somasundaram
CC TITLE OF INVENTION: THERAPEUTIC COMPOUNDS COMPRISED
CC TITLE OF INVENTION: OF ANTI-FC RECEPTOR ANTIBODIES
CC NUMBER OF SEQUENCES: 16
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: LAHIVE & COCKFIELD
CC STREET: 60 State Street, Suite 510
CC CITY: Boston
CC STATE: Massachusetts
CC COUNTRY: USA
CC ZIP: 02109-1875
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent in Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/661,052
CC FILING DATE:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/484,172
CC FILING DATE: 07-JUNE-1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Arnold, Beth E.
CC REGISTRATION NUMBER: 35,430
CC REFERENCE/DOCKET NUMBER: MXI-043CP
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617)227-7400
CC TELEFAX: (617)227-5941
CC INFORMATION FOR SEQ ID NO: 12:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 21 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC SEQUENCE 21 AA; 2479 MW; 2810 CN;

Query Match 100.0%; Score 158; DB 2; Length 21;
Best Local Similarity 100.0%; Pred. No. 3.25e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 FNNFTVSWLRVPKVSASHLE 21

QY 1 FNNFTVSWLRVPKVSASHLE 21

RESULT 4
ID PCT-US93-11703-66 STANDARD; PRT; 21 AA.

XX xxxxxx

Sequence 66, Application PC/TUS9311703

Sequence 66, Application PC/TUS9311703

GENERAL INFORMATION:

APPLICANT: Chiron Mimotopes Pty. Ltd.

TITLE OF INVENTION: T-Cell Epitopes

NUMBER OF SEQUENCES: 75

CORRESPONDENCE ADDRESS:

CC ADDRESSEE: Grant D. Green
CC STREET: 4560 Horton St.
CC CITY: Emeryville
CC STATE: CA
CC COUNTRY: USA
CC ZIP: 94508
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.30B
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US93/11703
CC FILING DATE: 28-DEC-1993
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/984,852
CC FILING DATE: 02-DEC-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Green, Grant D.
CC REGISTRATION NUMBER: 31,259
CC REFERENCE/DOCKET NUMBER: 0222.101
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 510-601-2706
CC TELEFAX: 510-655-3542
CC INFORMATION FOR SEQ ID NO: 66:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 21 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 21 AA; 2479 MW; 2810 CN;
Query Match 100.0%; Score 158; DB 3; Length 21;
Best Local Similarity 100.0%; Pred. No. 3.25e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 1 FNNFTVFWLRVFKVSASHLE 21
QY 1 FNNFTVFWLRVFKVSASHLE 21
RESULT 5
ID PCT-US93-11703-64 STANDARD; PRT; 31 AA.
XX
AC xxxxxx
DT
XX
DE Sequence 64, Application PC/TUS9311703
CC
CC Sequence 64, Application PC/TUS9311703
CC GENERAL INFORMATION:
CC APPLICANT: Chiron Mimotopes Pty. Ltd.
CC TITLE OF INVENTION: T-Cell Epitopes
CC NUMBER OF SEQUENCES: 75
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Grant D. Green
CC STREET: 4560 Horton St.
CC CITY: Emeryville
CC STATE: CA
CC COUNTRY: USA
CC ZIP: 94508
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.30B
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US93/11703
CC FILING DATE: 28-DEC-1993
CC CLASSIFICATION:

CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/984,852
CC FILING DATE: 02-DEC-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Green, Grant D.
CC REGISTRATION NUMBER: 31,259
CC REFERENCE/DOCKET NUMBER: 0222.101
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 510-601-2706
CC TELEFAX: 510-655-3542
CC INFORMATION FOR SEQ ID NO: 64:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 31 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 31 AA; 3694 MW; 6041 CN;
Query Match 100.0%; Score 158; DB 3; Length 31;
Best Local Similarity 100.0%; Pred. No. 3.25e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 7 FNNFTVFWLRVFKVSASHLE 27
QY 1 FNNFTVFWLRVFKVSASHLE 21
RESULT 6
ID US-08-488-351A-14 STANDARD; PRT; 32 AA.
XX
AC xxxxxx
DT
XX
DE Sequence 14, Application US/08488351A
XX
CC Sequence 14, Application US/08488351A
CC Patent No. 5843446
CC GENERAL INFORMATION:
CC APPLICANT: Ladd, Anna
CC APPLICANT: Wang, Chang Yi
CC APPLICANT: Zamb, Timothy
CC TITLE OF INVENTION: Immunogenic LHRH peptide constructs
CC TITLE OF INVENTION: and synthetic universal immune stimulators for vaccines
CC NUMBER OF SEQUENCES: 114
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Maria C.H. Lin
CC STREET: 345 Park Avenue
CC CITY: New York
CC STATE: NY
CC COUNTRY: US
CC ZIP: 10154-0053
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/488,351A
CC FILING DATE: 7-JUN-1995
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/446,692
CC FILING DATE: 7-JUN-1995
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/229,275
CC FILING DATE: 14-APR-1994
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/057,166
CC FILING DATE: 27-APR-1992

CC CLASSIFICATION: 424
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Maria C.H. Lin
CC REGISTRATION NUMBER: 29,323
CC REFERENCE/DOCKET NUMBER: 1151-4146 US2
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212)415-8745
CC TELEFAX: (516)751-6849
CC INFORMATION FOR SEQ ID NO: 14:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 32 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
SQ SEQUENCE 32 AA: 3789 MW; 6283 CN;

Query Match 100.0%; Score 158; DB 2; Length 32;
Best Local Similarity 100.0%; Pred.No. 3.25e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3 FNNFTVSFWLRVPKVSASHLE 23
QY 1 FNNFTVSFWLRVPKVSASHLE 21

RESULT 7
ID US-08-446-692-14 STANDARD; PRT; 32 AA.
XX xxxxxx
AC
XX
DT
DE
XX
Sequence 14, Application US/08446692
Sequence 14, Application US/08446692
Patent No. 5759551
GENERAL INFORMATION:
CC APPLICANT: Ladd, Anna
CC APPLICANT: Wang, Chang Yi
CC APPLICANT: Zamb, Timothy
CC TITLE OF INVENTION: Immunogenic LHRH peptide constructs
CC NUMBER OF SEQUENCES: 114
CC CORRESPONDENCE ADDRESSES:
CC ADDRESSER: Maria C.H. Lin
CC STREET: 345 Park Avenue
CC CITY: New York
CC STATE: NY
CC COUNTRY: US
CC ZIP: 10154-0053
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/446,692
CC FILING DATE: 7-JUN-1995
CC CLASSIFICATION: 424
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Maria C.H. Lin
CC REGISTRATION NUMBER: 29,323
CC REFERENCE/DOCKET NUMBER: 1151-4146 US2
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212)415-8745
CC TELEFAX: (516)751-6849
CC INFORMATION FOR SEQ ID NO: 14:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 32 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
SQ SEQUENCE 32 AA: 3789 MW; 6283 CN;

Query Match 100.0%; Score 158; DB 2; Length 32;
Best Local Similarity 100.0%; Pred.No. 3.25e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3 FNNFTVSFWLRVPKVSASHLE 23
QY 1 FNNFTVSFWLRVPKVSASHLE 21

RESULT 8
ID US-07-618-312A-4 STANDARD; PRT; 452 AA.
XX xxxxxx
AC
XX
DT
DE
XX
Sequence 4, Application US/07618312A
Sequence 4, Application US/07618312A
Patent No. 5389540
GENERAL INFORMATION:
CC APPLICANT: Makoff Dr, Andrew J
CC APPLICANT: Romanos Dr, Michael A
CC APPLICANT: Clare Dr, Jeffrey J
CC APPLICANT: Fairweather Dr, Neil F
CC TITLE OF INVENTION: VACCINES
CC NUMBER OF SEQUENCES: 13
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: 14th Floor
CC STREET: 2200 Clarendon Boulevard,
CC CITY: Arlington,
CC STATE: Virginia
CC COUNTRY: U.S.A.
CC ZIP: 22201
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/07/618,312A
CC FILING DATE: 19910516
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: GB 8926832.0
CC FILING DATE: 28-NOV-1989
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: GB 9006097.1
CC FILING DATE: 17-MAR-1990
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Crawford Mr, Arthur R
CC REGISTRATION NUMBER: 25,327
CC REFERENCE/DOCKET NUMBER: 510-51
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 0101 703 8750400
CC TELEFAX: 0101 703 5253468
CC TELE: 200797 NIXN UR
CC INFORMATION FOR SEQ ID NO: 4:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 452 amino acids
CC TYPE: AMINO ACID
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
SQ SEQUENCE 452 AA; 51770 MW; 1065961 CN;

Query Match 100.0%; Score 158; DB 1; Length 452;
Best Local Similarity 100.0%; Pred.No. 3.25e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 84 FNNFTVSFWLRVPKVSASHLE 104
QY 1 FNNFTVSFWLRVPKVSASHLE 21


```
CC Patent No. 5571694
CC GENERAL INFORMATION:
CC APPLICANT: Makoff Dr, Andrew J
CC APPLICANT: Romanos Dr, Michael A
CC APPLICANT: Claire Dr, Jeffrey J
CC APPLICANT: Fairweather Dr, Neil F
CC TITLE OF INVENTION: VACCINES
CC NUMBER OF SEQUENCES: 13
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: NIXON & VANDERHYE P.C.
CC STREET: 1100 No. 5571694th Glebe Road
CC CITY: Arlington,
CC STATE: Virginia
CC COUNTRY: U.S.A.
CC ZIP: 22201-4714
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/280,228
CC FILING DATE: 25-JUL-1994
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC PRIOR APPLICATION NUMBER: US 07/618,312
CC FILING DATE: 27-NOV-1990
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: GB 8526832.0
CC FILING DATE: 28-NOV-1989
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: GB 9006097.1
CC FILING DATE: 17-MAR-1990
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Wilson, Mary J.
CC REGISTRATION NUMBER: 32,955
CC REFERENCE/DOCKET NUMBER: 117-163
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (703) 816-4000
CC TELEFAX: (703) 816-4100
CC TELEFAX: 200797 NIXN UR
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 452 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 452 AA; 51797 MW; 1065526 CN;

Query Match 100.0%; Score 158; DB 1; Length 452;
Best Local Similarity 100.0%; Pred. No. 3.25e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 84 FNNFTVSFWLRVPKVSASHLE 104
QY 1 FNNFTVSFWLRVPKVSASHLE 21

RESULT 12
ID US-08-110-786A-8 STANDARD; PRT; 452 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 8, Application US/08110786A
XX
DE Patent No. 5443966
CC
CC TITLE OF INVENTION:
CC GENERAL INFORMATION:
CC APPLICANT: FAIRWEATHER, Neil Fraser
CC APPLICANT: MAKOFF, Andrew Joseph

CC Patent No. 5571694
CC GENERAL INFORMATION:
CC APPLICANT: Makoff Dr, Andrew J
CC APPLICANT: Romanos Dr, Michael A
CC APPLICANT: Claire Dr, Jeffrey J
CC APPLICANT: Fairweather Dr, Neil F
CC TITLE OF INVENTION: VACCINES
CC NUMBER OF SEQUENCES: 13
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: NIXON & VANDERHYE P.C.
CC STREET: 1100 No. 5571694th Glebe Road
CC CITY: Arlington,
CC STATE: Virginia
CC COUNTRY: U.S.A.
CC ZIP: 22201-4714
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/280,228
CC FILING DATE: 25-JUL-1994
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC PRIOR APPLICATION NUMBER: US 07/618,312
CC FILING DATE: 27-NOV-1990
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: GB 8526832.0
CC FILING DATE: 28-NOV-1989
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: GB 9006097.1
CC FILING DATE: 17-MAR-1990
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Wilson, Mary J.
CC REGISTRATION NUMBER: 32,955
CC REFERENCE/DOCKET NUMBER: 117-163
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (703) 816-4000
CC TELEFAX: (703) 816-4100
CC TELEFAX: 200797 NIXN UR
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 452 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 452 AA; 51797 MW; 1065526 CN;

Query Match 100.0%; Score 158; DB 1; Length 452;
Best Local Similarity 100.0%; Pred. No. 3.25e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 84 FNNFTVSFWLRVPKVSASHLE 104
QY 1 FNNFTVSFWLRVPKVSASHLE 21

RESULT 13
ID US-08-668-381A-5 STANDARD; PRT; 618 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 5, Application US/08668381A
XX
DE Patent No. 5780024
CC
CC GENERAL INFORMATION:
CC APPLICANT: Brown, Robert H.
CC APPLICANT: Fishman, Paul S.
CC APPLICANT: Francis, Jonathan W.
CC APPLICANT: Hosler, Betsy A.
CC TITLE OF INVENTION: SUPEROXIDE DISMUTASE/TETANUS TOXIN
CC TITLE OF INVENTION: FRAGMENT C HYBRID PROTEIN
CC NUMBER OF SEQUENCES: 6
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Fish & Richardson P.C.
```

```
CC STREET: 225 Franklin Street
CC CITY: Boston
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02110-2804
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/668,381A
CC FILING DATE: 21-JUN-1996
CC CLASSIFICATION: 514
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 60/000,473
CC FILING DATE: 23-JUN-1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Clark, Paul T.
CC REGISTRATION NUMBER: 30,164
CC REFERENCE/DOCKET NUMBER: 00786/269001
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 617/542-5070
CC TELEFAX: 617/542-8906
CC TELEX: 200154
CC INFORMATION FOR SEQ ID NO: 5:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 618 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 618 AA: 68895 MW; 1991829 CN;
Query Match 100.0%; Score 158; DB 2; Length 618;
Best Local Similarity 100.0%; Pred. No. 3.25e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 250 FNNFTVSFWLRVPKVSASHLE 270
QY 1 FNNFTVSFWLRVPKVSASHLE 21
RESULT 14
ID PCT-US95-13841-8 STANDARD; PRT; 22 AA.
XX xxxxxx
AC
XX
XX
DT
XX
DE
XX
XX
CC Sequence 8, Application PC/TUS9513841
CC GENERAL INFORMATION:
CC APPLICANT: United Biomedical Inc; Walfield, Alan M.;
CC APPLICANT: Wang, Chang Yi
CC TITLE OF INVENTION: Synthetic IgE Membrane Anchor
CC TITLE OF INVENTION: Peptide Immunogens for the Treatment of Allergy
CC NUMBER OF SEQUENCES: 32
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Maria C.H. Lin
CC STREET: 345 Park Avenue
CC CITY: New York
CC STATE: NY
CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Wordperfect 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/13841
CC FILING DATE: 25-OCT-1995
CC STREET: 225 Franklin Street
CC CITY: Boston
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02110-2804
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/668,381A
CC FILING DATE: 21-JUN-1996
CC CLASSIFICATION: 514
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 60/000,473
CC FILING DATE: 23-JUN-1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Clark, Paul T.
CC REGISTRATION NUMBER: 30,164
CC REFERENCE/DOCKET NUMBER: 00786/269001
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 617/542-5070
CC TELEFAX: 617/542-8906
CC TELEX: 200154
CC INFORMATION FOR SEQ ID NO: 5:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 618 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 618 AA: 68895 MW; 1991829 CN;
Query Match 100.0%; Score 158; DB 2; Length 618;
Best Local Similarity 100.0%; Pred. No. 3.25e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 250 FNNFTVSFWLRVPKVSASHLE 270
QY 1 FNNFTVSFWLRVPKVSASHLE 21
RESULT 14
ID PCT-US95-13841-8 STANDARD; PRT; 22 AA.
XX xxxxxx
AC
XX
XX
DT
XX
DE
XX
XX
CC Sequence 8, Application PC/TUS9513841
CC GENERAL INFORMATION:
CC APPLICANT: United Biomedical Inc; Walfield, Alan M.;
CC APPLICANT: Wang, Chang Yi
CC TITLE OF INVENTION: Synthetic IgE Membrane Anchor
CC TITLE OF INVENTION: Peptide Immunogens for the Treatment of Allergy
CC NUMBER OF SEQUENCES: 32
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Maria C.H. Lin
CC STREET: 345 Park Avenue
CC CITY: New York
CC STATE: NY
CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Wordperfect 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/13841
CC FILING DATE: 25-OCT-1995
CC STREET: 225 Franklin Street
CC CITY: Boston
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02110-2804
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/668,381A
CC FILING DATE: 21-JUN-1996
CC CLASSIFICATION: 514
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 60/000,473
CC FILING DATE: 23-JUN-1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Lin, Maria C.H.
CC REGISTRATION NUMBER: 29,323
CC REFERENCE/DOCKET NUMBER: 1151-4117
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 212-758-4800
CC TELEFAX: 212-751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 8:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 22 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 22 AA: 2606 MW; 3203 CN;
Query Match 96.2%; Score 152; DB 3; Length 22;
Best Local Similarity 100.0%; Pred. No. 1.50e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 3 FNNFTVSFWLRVPKVSASHL 22
QY 1 FNNFTVSFWLRVPKVSASHL 20
RESULT 15
ID US-08-488-351A-5 STANDARD; PRT; 22 AA.
XX xxxxxx
AC
XX
XX
DT
XX
DE
XX
XX
CC Sequence 5, Application US/08488351A
CC Patent No. 5843446
CC GENERAL INFORMATION:
CC APPLICANT: Ladd, Anna
CC APPLICANT: Wang, Chang Yi
CC APPLICANT: Zamb, Timothy
CC TITLE OF INVENTION: Immunogenic LHRH peptide constructs
CC TITLE OF INVENTION: and synthetic universal immune stimulators for vaccines
CC NUMBER OF SEQUENCES: 114
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Maria C.H. Lin
CC STREET: 345 Park Avenue
CC CITY: New York
CC STATE: NY
CC COUNTRY: US
CC ZIP: 10154-0053
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/488,351A
CC FILING DATE: 7-JUN-1995
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/446,692
CC FILING DATE: 7-JUN-1995
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/229,275
CC FILING DATE: 14-APR-1994
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
```

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Tue Aug 17 15:58:29 1999; MasPar time 4.18 Seconds
 Tabular output not generated.
 114.914 Million cell updates/sec

Title: >US-09-049-847-3
 Description: (1-12) from US09049847.ppe
 Perfect Score: 82
 Sequence: 1 GOICNDPNRDL 12

Scoring table: PAM 150
 Gap 15
 Searched: 122810 seqs, 40068593 residues
 Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: pir60
 1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 23.828; Variance 28.469; scale 0.837

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES						
Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	82	100.0	1315	1	tentoxylisin (EC 3.4.1.1)	1.58e+07
2	52	63.4	403	2	probable transglutaminase	2.85e+00
3	52	63.4	428	2	26S proteasome regulin	2.85e+00
4	52	63.4	3411	1	genome polyprotein -	2.85e+00
5	52	63.4	3411	1	genome polyprotein -	2.85e+00
6	51	62.2	190	2	GTP cyclohydrolase I	4.61e+00
7	51	62.2	190	2	GTP cyclohydrolase I	4.61e+00
8	51	62.2	364	1	matrix protein - Nesc	4.61e+00
9	51	62.2	548	2	glucan endo-1,3-beta-	4.61e+00
10	51	62.2	1115	2	PAN2 protein - yeast	4.61e+00
11	51	62.2	3163	1	genome polyprotein -	4.61e+00
12	50	61.0	127	2	aspartate 1-decarboxylase	7.43e+00
13	50	61.0	202	2	GTP cyclohydrolase I	7.43e+00
14	50	61.0	522	1	protein disulfide-iso	7.43e+00
15	50	61.0	633	2	BEA4 protein - yeast	7.43e+00
16	50	61.0	3770	2	delta-(L-alpha-amino)	7.43e+00
17	49	59.8	184	2	GTP cyclohydrolase I	1.19e+01
18	49	59.8	841	2	protein-tyrosine-phos	1.19e+01
19	49	59.8	1346	2	probable regulatory p	1.19e+01
20	48	58.5	73	2	protein YNR032c-a	1.88e+01
21	48	58.5	144	2	methylmalonyl-CoA mut	1.88e+01
22	48	58.5	179	2	GTP cyclohydrolase I	1.88e+01
23	48	58.5	261	2	hypothetical protein	1.88e+01

24 48 58.5 296 2 S75858 glutamine-binding pro 1.88e+01
 25 48 58.5 509 1 DENIN4 NADH dehydrogenase (u 1.88e+01
 26 48 58.5 802 1 A26343 ADE5 multifunctional 1.88e+01
 27 47 57.3 84 1 Q1BFS2 uracil-DNA glycosylas 2.97e+01
 28 47 57.3 299 2 B70951 hypothetical protein 2.97e+01
 29 47 57.3 378 2 B38178 teta protein - plasm 2.97e+01
 30 47 57.3 514 2 A35658 transcription factor 2.97e+01
 31 47 57.3 562 2 H70888 hypothetical protein 2.97e+01
 32 47 57.3 680 2 A28121 major merozoite surfa 2.97e+01
 33 47 57.3 796 1 JV0107 glucose dehydrogenase 2.97e+01
 34 47 57.3 821 1 A39616 protein kinase RAD53 2.97e+01
 35 47 57.3 1379 2 S01254 hepatocyte growth fac 2.97e+01
 36 47 57.3 1772 2 A45532 conserved hypothetical 4.64e+01
 37 46 56.1 250 2 D69182 conserved hypothetical 4.64e+01
 38 46 56.1 282 2 C69553 gene 9 protein - phag 4.64e+01
 39 46 56.1 288 1 GNBPT4 hypothetical protein 4.64e+01
 40 46 56.1 320 2 D71098 tryptophan synthase (4.64e+01
 41 46 56.1 397 2 I64122 cyclomaltodextrin glu 4.64e+01
 42 46 56.1 655 1 ALKBG collagen alpha 1(I) c 4.64e+01
 43 46 56.1 791 1 CG901S kinesin-related prote 4.64e+01
 44 46 56.1 1184 1 A34795 zinc-finger protein - 4.64e+01
 45 46 56.1 2180 2 A47651

ALIGNMENTS

RESULT 1
 ENTRY BTCLIN #type complete
 TITLE tentoxylisin (EC 3.4.24.68) precursor - Clostridium tetani
 ALTERNATE_NAMES tetanus neurotoxin
 ORGANISM #formal_name Clostridium tetani
 DATE 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 26-Feb-1999
 ACCESSIONS A25689; A25757; A25194; B25194; A60759; S69348; S09364
 REFERENCE A25689
 #authors Eisel, U.; Jarausch, W.; Goretzki, K.; Henschen, A.; Engels, J.; Weller, U.; Hudel, M.; Habermann, E.; Niemann, R.
 #journal EMBO J. (1986) 5:2495-2502
 #title Tetanus toxin: primary structure, expression in E. coli, and homology with botulinum toxins.
 #cross-references MUID:87053814
 #accession A25689
 #molecule_type DNA
 #residues 1-1315 #label EIS
 #cross-references GB:X04436; NID:g40769; PID:g40770
 REFERENCE A25757
 #authors Fairweather, N.F.; Lyness, V.A.
 #journal Nucleic Acids Res. (1986) 14:7809-7812
 #title The complete nucleotide sequence of tetanus toxin.
 #cross-references MUID:87040747
 #accession A25757
 #molecule_type DNA
 #residues 1-1315 #label FAI
 #cross-references GB:X06214; NID:g40773; PID:g40774
 #experimental_source strain CN3911
 REFERENCE A25194
 #authors Fairweather, N.F.; Lyness, V.A.; Pickard, D.J.; Allen, G.; Thomson, R.O.
 #journal J. Bacteriol. (1986) 165:21-27
 #title Cloning, nucleotide sequencing, and expression of tetanus toxin fragment C in Escherichia coli.
 #cross-references MUID:86085672
 #accession A25194
 #molecule_type DNA
 #residues 743-1315 #label FA2
 #cross-references GB:M12739; NID:g144920; PID:g144921
 #accession B25194
 #molecule_type protein
 #residues 865-894 #label FA3
 REFERENCE A60759
 #authors Matsuda, M.; Lei, D.L.; Sugimoto, N.; Ozutsumi, K.; Okabe, T.
 #journal Infect. Immun. (1989) 57:3588-3593
 #title Isolation, purification, and characterization of fragment B,

the NH-2-terminal half of the heavy chain of tetanus toxin.

```
#cross-references MUID:90035436
#accession A60759
#molecule_type protein
#residues 461-475 ##label MAT
REFERENCE JS0098
#authors Demotz, S.; Lanzavecchia, L.; Eisel, U.; Niemann, H.;
#journal Wildmann, C.; Corradin, G.
#title J. Immunol. (1989) 142:394-402
#text_change Delineation of several DR-restricted tetanus toxin T cell
epitopes.
#cross-references MUID:89093918
#contents annotation: epitope region
REFERENCE S27125
#authors Schiavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de
#journal Lauroto, P.P.; Dasgupta, B.R.; Montecucco, C.
#title Nature (1992) 359:832-835
#text_change Tetanus and botulinum-B neurotoxins block neurotransmitter
release by proteolytic cleavage of synaptobrevin.
#cross-references MUID:93063293
#contents annotation
REFERENCE S69348
#authors de Filippis, V.; Vangelista, L.; Schiavo, G.; Tonello, F.;
#journal Montecucco, C.
#title Eur. J. Biochem. (1995) 229:61-69
#text_change Structural studies on the zinc-endopeptidase light chain of
tetanus neurotoxin.
#cross-references MUID:95262688
#accession S69348
#molecule_type protein
#residues 2-31 ##label DEF
COMMENT The source of this protein was an extrachromosomal plasmid.
COMMENT The precursor is cleaved by endogenous proteinase activity to form
light (fragment A) and heavy (fragment B.C) chains that are
covalently linked by an interchain disulfide bond (the individual
chains are not toxic when separated). The amino end of the heavy
chain (fragment B) can be separated from the carboxyl end
(fragment C) by papain.
COMMENT Fragment B forms ion channels in a lipid bilayer. Fragment C binds
to gangliosides and may target the toxin to the motor end plate.
COMMENT This potent neurotoxin binds to peripheral neuronal synapses, is
internalized, and moves by retrograde transport up the axon into
the spinal cord where it can move between postsynaptic and
presynaptic neurons. It inhibits neurotransmitter release by
proteolytic cleavage of synaptobrevin (vesicle-associated
membrane protein 2).
FUNCTION
#description blocks neuroexocytosis via hydrolysis of a Gln-Phe peptide
bond in synaptobrevin 2
CLASSIFICATION #superfamily tetanus toxin
KEYWORDS hydrolase; metalloproteinase; neurotoxin; transmembrane
protein; zinc
FEATURE
2-457
#product tentoxilysin light chain (fragment A) #status
predicted #label TTL\
461-1315 #product tentoxilysin heavy chain (fragment B.C) #status
experimental #label TTH\
461-864 #domain channel forming (fragment B) #status predicted
#label TxB\
865-1315 #domain ganglioside binding (fragment C) #status
predicted #label TxC\
233,237 #binding_site zinc (His) #status predicted\
234 #active_site glu #status predicted
SUMMARY #length 1315 #molecular-weight 150681 #checksum 4853
Query Match 100.0%; Score 82; DB 1; Length 1315;
Best Local Similarity 100.0%; Pred. No. 1.56e-07;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 1273 GOIGNDPNRDIL 1284
|||||
QY 1 GOIGNDPNRDIL 12
```

RESULT 2

ENTRY
TITLE

JN0611 #type complete
probable transcription factor DdTPP2 - slime mold
(Dictyostelium discoideum)

ALTERNATE_NAMES
ORGANISM

Tat-binding protein 2
#formal_name Dictyostelium discoideum
31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change
04-Sep-1998

ACCESSIONS
REFERENCE

JN0611
JN0610
Shaw, D.R.; Ennis, H.L.
Biochem. Biophys. Res. Commun. (1993) 193:1291-1296

#authors
#journal
#title

Molecular cloning and developmental regulation of
Dictyostelium discoideum homologues of the human and yeast
HIV1 Tat-binding protein.

#cross-references MUID:93312322

#accession
##status

JN0611
nucleic acid sequence not shown

##molecule_type mRNA
##residues 1-403 ##label SHA

#cross-references GB:LI6578; NID:g290054; PID:g290055

CLASSIFICATION

#superfamily ATP-dependent 26S proteinase;
FtsH/SEC18/CDC48-type ATP-binding domain homology
ATP; nucleus; P-loop

KEYWORDS
FEATURE

164-374

#domain FtsH/SEC18/CDC48-type ATP-binding domain
homology #label VATP\

191-198

#region nucleotide-binding motif A (P-loop)
#length 403 #molecular-weight 45542 #checksum 8766

SUMMARY

Query Match 63.4%; Score 52; DB 2; Length 403;
Best Local Similarity 63.6%; Pred. No. 2.85e+00;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 178 QIGIDPPRGVL 188
|||||

QY 2 QIGNDPNRDIL 12

RESULT 3

ENTRY
TITLE

S69678 #type complete
26S proteasome regulatory particle chain RPT3 - yeast
(Saccharomyces cerevisiae)

ALTERNATE_NAMES

protein YDR394W; YTA2 protein

ORGANISM

#formal_name Saccharomyces cerevisiae
22-Aug-1996 #sequence_revision 06-Sep-1996 #text_change

DATE

13-Sep-1998

ACCESSIONS
REFERENCE

S69678; S46606; S34353
S69665

#authors
#submission

Dietrich, F.S.
submitted to the EMBL Data Library, July 1995

#description

The sequence of S. cerevisiae cosmid 9481, 9509, 9926, 9461,
and lambda 3641.

#accession
##molecule_type DNA

S69678
#residues 1-428 ##label DIE

##cross-references

EMBL:U32274; NID:g927313; PID:g927327; MIPS:YDR394W

REFERENCE

S46605

#authors

Schnall, R.; Mannhaupt, G.; Stucka, R.; Tauer, R.; Ehnlé, S.;
Schwarzlose, C.; Vetter, I.; Felkmann, H.

#journal

Yeast (1994) 10:1141-1155

#title

Identification of a set of yeast genes coding for a novel
family of putative ATPases with high similarity to
constituents of the 26S protease complex.

#cross-references MUID:95274317

#accession S46606

##status

nucleic acid sequence not shown

##molecule_type DNA

#residues 1-341,'Y',343-428 ##label SCH

##cross-references

EMBL:X73570; NID:g313879; PID:g313880

GENETICS

#gene SGD:RPT3; YTA2; YN11
 ##cross-references SGD:S0002802; MIPS:YDR394w
 #map_position 4R
 CLASSIFICATION #superfamily ATP-dependent 26S proteinase;
 FtsH/SEC18/CDC48-type ATP-binding domain homology
 nucleus; P-loop
 KEYWORDS
 FEATURE
 186-397 #domain FtsH/SEC18/CDC48-type ATP-binding domain
 homology #label VARP
 213-220 #region nucleotide-binding motif A (P-loop)
 SUMMARY
 #length 428 #molecular-weight 47893 #checksum 5711

Query Match 63.48; Score 52; DB 2; Length 428;
 Best Local Similarity 63.68; Pred.No. 2.85e+00;
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 200 QIGIDPPRGVL 210
 ||| || | |
 QY 2 QIGNDPNRDL 12

RESULT 4
 ENTRY #type complete
 TITLE genome polyprotein - yellow fever virus (strain 17D)
 CONTAINS capsid protein C; envelope protein M; major envelope protein E; nonstructural protein NS1; nonstructural protein NS2a; nonstructural protein NS2b; nonstructural protein NS3; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
 ORGANISM #formal_name yellow fever virus
 DATE 27-Nov-1985 #sequence_revision 27-Nov-1985 #text_change 15-May-1998
 A03914
 REFERENCES
 #authors Rice, C.M.; Lenches, E.M.; Eddy, S.R.; Shin, S.J.; Sheets, R.L.; Strauss, J.H.
 #journal Science (1985) 229:726-733
 #title Nucleotide sequence of yellow fever virus: implications for flavivirus gene expression and evolution.
 ##cross-references MUID:85272570
 #accession A03914
 ##molecule_type genomic RNA
 ##residues 1-3411 ##label RIC
 CLASSIFICATION #superfamily yellow fever virus genome polyprotein; DEAD/H box helicase homology
 KEYWORDS ATP; capsid protein; envelope protein; glycoprotein; nonstructural protein; P-loop; polyprotein; transmembrane protein

FEATURE
 2-210 #product capsid protein C #status predicted #label CPC
 211-285 #product envelope protein M #status predicted #label EPM
 249-269 #domain transmembrane #status predicted #label TM1
 271-285 #domain transmembrane #status predicted #label TM2
 286-778 #product major envelope protein E #status predicted #label MEE
 740-753 #domain transmembrane #status predicted #label TM3
 755-778 #domain transmembrane #status predicted #label TM4
 779-1187 #product nonstructural protein NS1 #status predicted #label NS1
 1159-1180 #domain transmembrane #status predicted #label TM5
 1188-1354 #product nonstructural protein NS2a #status predicted #label N2A
 1355-1484 #product nonstructural protein NS2b #status predicted #label N2B
 1485-2107 #product nonstructural protein NS3 #status predicted #label NS3
 1682-1951 #domain DEAD/H box helicase homology #label DEAD
 1682-1689 #region nucleotide-binding motif A (P-loop)
 1769-1774 #region nucleotide-binding motif B
 1773-1776 #region DEAD motif
 2108-2394 #product nonstructural protein NS4a #status predicted #label N4A

2395-2506 #product nonstructural protein NS4b #status predicted #label N4B
 2507-3411 #product nonstructural protein NS5 #status predicted #label NS5
 134,150,172,266,
 594,755,908,986,
 1796,2062,2320,
 2346,2408,2467,
 2720,2734,2740
 #binding_site carbohydrate (Asn) (covalent) #status predicted
 SUMMARY
 #length 3411 #molecular-weight 379516 #checksum 8719
 Query Match 63.48; Score 52; DB 1; Length 3411;
 Best Local Similarity 70.08; Pred.No. 2.85e+00;
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 1947 GRIGRPNRD 1956
 | : | | : | | |
 QY 1 QGIGNDPNRD 10

RESULT 5
 ENTRY #type complete
 TITLE genome polyprotein - yellow fever virus (strain Pasteur 17D-204)
 CONTAINS capsid protein C; envelope protein M; major envelope protein E; nonstructural protein NS1; nonstructural protein NS2a; nonstructural protein NS2b; nonstructural protein NS3; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
 ORGANISM #formal_name yellow fever virus
 DATE 31-Mar-1991 #sequence_revision 31-Mar-1991 #text_change 02-Jul-1998
 S07757
 REFERENCES
 #authors Dupuy, A.; Despres, P.; Cahour, A.; Girard, M.; Bouloy, M.
 #journal Nucleic Acids Res. (1989) 17:3989
 #title Nucleotide sequence comparison of the genome of two 17D-204 yellow fever vaccines.
 ##cross-references MUID:89282413
 #accession S07757
 ##status nucleic acid sequence not shown; translation not shown
 ##molecule_type genomic RNA
 ##residues 1-3411 ##label DUP
 ##cross-references EMBL:X15062
 ##note the nucleotide sequence was submitted to the EMBL Data Library, April 1989, in computer-readable form

CLASSIFICATION #superfamily yellow fever virus genome polyprotein; DEAD/H box helicase homology
 KEYWORDS ATP; capsid protein; envelope protein; glycoprotein; nonstructural protein; P-loop; polyprotein; transmembrane protein
 FEATURE
 2-210 #product capsid protein C #status predicted #label CPC
 105-125 #domain transmembrane #status predicted #label TM1
 211-285 #product envelope protein M #status predicted #label EPM
 271-289 #domain transmembrane #status predicted #label TM2
 286-778 #product major envelope protein E #status predicted #label MEE
 736-753 #domain transmembrane #status predicted #label TM3
 756-778 #domain transmembrane #status predicted #label TM4
 779-1187 #product nonstructural protein NS1 #status predicted #label NS1
 1133-1151 #domain transmembrane #status predicted #label TM5
 1160-1179 #domain transmembrane #status predicted #label TM6
 1188-1354 #product nonstructural protein NS2a #status predicted #label N2A
 1355-1484 #product nonstructural protein NS2b #status predicted #label N2B
 1485-2107 #product nonstructural protein NS3 #status predicted #label NS3
 1682-1951 #domain DEAD/H box helicase homology #label DEAD

```

1682-1689      #region nucleotide-binding motif A (P-loop)\
1769-1774      #region nucleotide-binding motif B\
1773-1776      #region DEAH motif\
2108-2394      #product nonstructural protein NS4a #status predicted
               #label N4A\
2395-2506      #product nonstructural protein NS4b #status predicted
               #label N4B\
2507-3411      #product nonstructural protein NS5 #status predicted
               #label NS5\
134,150,172,594,
908,986,1796,2062,
2320,2346,2408,
2467,2720,2734,
2740           #binding_site carbohydrate (Asn) (covalent) #status
               predicted
SUMMARY        #length 3411 #molecular-weight 379528 #checksum 8417
               #length 3411 #molecular-weight 379528 #checksum 8417
Query Match    63.4%; Score 52; DB 1; Length 3411;
Best Local Similarity 70.0%; Pred. No. 2.85e+00;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 1947 GRIGNPRD 1956
QY 1 GOIGNDPRD 10
1:|||||
2:|||||

RESULT 6
ENTRY   F71695      #type complete
TITLE   GTP cyclohydrolase I (foIE) RP383 - Rickettsia prowazekii
ORGANISM #formal_name Rickettsia prowazekii
DATE    21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change
21-Nov-1998
ACCESSIONS F71695
REFERENCE  A71630
#authors   Andersson, S.G.E.; Zomrodipour, A.; Andersson, J.O.;
           Sacheritz-Ponten, T.; Alsmark, U.C.M.; Podowski, R.M.;
           Naeslund, A.K.; Eriksson, A.S.; Winkler, H.H.; Kurland,
           C.G.
#journal   Nature (1998) 396:133-140
#title     The genome sequence of Rickettsia prowazekii and the origin
           of mitochondria.
#accession F71695
#status    preliminary; nucleic acid sequence not shown;
           translation not shown
#molecule_type DNA
#residues  1-190 #label AND
#cross-references GB:AJ235269; GB:AJ235269; NID:g3860788; PID:e1342684;
#experimental_source strain Madrid E

GENETICS
#gene      foIE; RP383
SUMMARY    #length 190 #molecular-weight 21790 #checksum 3341

Query Match    52.2%; Score 51; DB 2; Length 190;
Best Local Similarity 70.0%; Pred. No. 4.61e+00;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 20 IGSDPNRGL 29
QY 3 IGNDPNRDL 12
1:|||||
2:|||||

RESULT 7
ENTRY   A38256      #type complete
TITLE   GTP cyclohydrolase I (EC 3.5.4.16) mtrA - Bacillus subtilis
ORGANISM #formal_name Bacillus subtilis
DATE    14-Jun-1991 #sequence_revision 14-Jun-1991 #text_change
24-Sep-1998
ACCESSIONS A38256; A39409; G69661
REFERENCE  A38256
#authors   Gollnick, P.; Ishino, S.; Kuroda, M.I.; Henner, D.J.;
           Yanofsky, C.
#journal   Proc. Natl. Acad. Sci. U.S.A. (1990) 87:8726-8730

```

```

#title      The mtr locus is a two-gene operon required for transcription
           attenuation in the trp operon of Bacillus subtilis.
#cross-references MUID:91062353
#accession  A38256
#molecule_type DNA
#residues   1-190 #label GOL
#cross-references GB:M37320; NID:gl43230; PID:gl43231
REFERENCE  A39409
#authors    Micka, B.; Groch, N.; Heinemann, U.; Marahiel, M.A.
           J. Bacteriol. (1991) 173:3191-3198
#title      Molecular cloning, nucleotide sequence, and characterization
           of the Bacillus subtilis gene encoding the DNA-binding
           protein HBSu.
#cross-references MUID:91216992
#accession  A39409
#molecule_type DNA
#residues   1-19 #label MIC
#cross-references GB:X52418
REFERENCE  A69380
#authors    Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
           Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
           Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
           A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
           Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
           Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
           Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.;
           Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.;
           Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
           M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
           S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
           Gulseppi, G.; Guy, B.J.; Haga, K.; Halesch, J.; Harwood,
           C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
           Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
           Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
           Y.; Koettler, P.; Konigstein, G.; Krogh, S.; Kumano, M.;
           Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
           Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
           Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
           M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
           M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,
           V.; Pohl, T.M.; Portetel, D.; Porwollik, S.; Prescott,
           A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
           Rey, M.; Reynolds, S.; Riéger, M.; Rivoita, C.; Rocha, E.;
           Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.;
           Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
           Sekowska, A.; Seror, S.J.; Serro, P.; Shin, B.S.; Soldo,
           B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
           Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
           Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
           Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
           Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.;
           Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
           K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
           Yoshikawa, H.; Panchin, A.
           Nature (1997) 390:249-256
#journal    The complete genome sequence of the Gram-positive bacterium
           Bacillus subtilis.
#title      #cross-references MUID:98044033
#accession  G69661
#status     nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues   1-190 #label KUN
#cross-references GB:Z99115; GB:AL009126; NID:g2634478; PID:e1183723;
           PID:g2634696
#experimental_source strain 168
COMMENT     This enzyme catalyzes the first step of the synthesis of
           tetrahydrobiopterin, a cofactor in the synthesis of aromatic
           amino acids.
GENETICS
#gene      mtrA
#note      the two genes in this methyltryptophan resistance (mtr)
           operon are mtrA and mtrB
#classification #superfamily GTP cyclohydrolase I

```

```

KEYWORDS
SUMMARY
  hydrolase
  #length 190 #molecular-weight 21219 #checksum 1098

Query Match      62.2%; Score 51; DB 2; Length 190;
Best Local Similarity 70.0%; Pred. No. 4.61e+00;
Matches          7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 20 IGDPNREGI 29
QY 3 IGDPNREGI 12

RESULT 8
ENTRY MFVZNC #type complete
TITLE matrix protein - Newcastle disease virus (strain Beaudette C)
ALTERNATE_NAMES #formal_name Newcastle disease virus
ORGANISM 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change
DATE 20-Mar-1998
ACCESSIONS A26111
REFERENCE A26111
#authors Chambers, P.; Millar, N.S.; Platt, S.G.; Emmerson, P.T.
#journal Nucleic Acids Res. (1986) 14:9051-9061
#title Nucleotide sequence of the gene encoding the matrix protein
#cross-references MUID:87066775
#accession A26111
##molecule_type genomic RNA
##residues 1-364 #label CHA
##cross-references GB:X04687; NID:g60940; PID:g60941

GENETICS
#gene M
CLASSIFICATION #superfamily parainfluenza virus matrix protein
KEYWORDS matrix protein
SUMMARY #length 364 #molecular-weight 39604 #checksum 4208

Query Match      62.2%; Score 51; DB 1; Length 364;
Best Local Similarity 41.7%; Pred. No. 4.61e+00;
Matches          5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 76 GMINDPKRELL 87
QY 1 GQIGNDPNRDI 12

RESULT 9
ENTRY A39094 #type complete
TITLE glucan endo-1,3-beta-glucosidase (EC 3.2.1.-) precursor -
  Oerskovia xanthineolytica
ORGANISM #formal_name Oerskovia xanthineolytica
DATE 27-Nov-1991 #sequence_revision 27-Nov-1991 #text_change
ACCESSIONS A39094
REFERENCE A39094
#authors Shen, S.H.; Chretien, P.; Bastien, L.; Slliaty, S.N.
#journal J. Biol. Chem. (1991) 266:1058-1063
#title Primary sequence of the glucanase gene from Oerskovia
  xanthineolytica. Expression and purification of the enzyme
  from Escherichia coli.
#cross-references MUID:91093212
#accession A39094
##status Preliminary
##molecule_type DNA
##residues 1-548 #label SHE
##cross-references GB:M60826; GB:M38734; NID:g150444; PID:g150445
KEYWORDS glucosidase; hydrolase
SUMMARY #length 548 #molecular-weight 58088 #checksum 8536

Query Match      62.2%; Score 51; DB 2; Length 548;
Best Local Similarity 53.6%; Pred. No. 4.61e+00;
Matches          7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 138 QNDSDFNRNII 148
QY 1 QNDSDFNRNII 148

```

```

QY 2 QIGNDPNRDI 12

RESULT 10
ENTRY S64101 #type complete
TITLE PAN2 protein - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES protein G3165; protein YGL094c
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 17-May-1996 #sequence_revision 17-May-1996 #text_change
  06-Feb-1998
ACCESSIONS S64101
REFERENCE S64071
#authors Rieger, M.; Mueller-Auer, S.; Brueckner, M.; Schaefer, M.
#submission submitted to the Protein Sequence Database, May 1996
#accession S64101
##molecule_type DNA
##residues 1-1115 #label RIE
##cross-references EMBL:Z72516; NID:g1322626; PID:e243300; PID:g1322627;
  MIPS:YGL094c
##experimental_source strain S288C
GENETICS
#gene SGD:PAN2
CLASSIFICATION #cross-references SGD:S0003062; MIPS:YGL094c
KEYWORDS #map_position 7L
SUMMARY #length 1115 #molecular-weight 127038 #checksum 4836

Query Match      62.2%; Score 51; DB 2; Length 1115;
Best Local Similarity 50.0%; Pred. No. 4.61e+00;
Matches          6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 68 GHIGGNSVKDIL 79
QY 1 GQIGNDPNRDI 12

RESULT 11
ENTRY JQ1895 #type complete
TITLE genome polyprotein - turnip mosaic virus (strain Quebec)
CONTAINS coat protein; cytoplasmic inclusion protein; helper component
  protein; nuclear inclusion a protein; nuclear inclusion b
  protein; P1 protein; P3 protein; p6K1 protein; p6K2
  protein; VPg protein
ORGANISM #formal_name turnip mosaic virus, TuMV
DATE 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change
ACCESSIONS JQ1895
REFERENCE JQ1895
#authors Nicolas, O.; Laliberte, J.F.
#journal J. Gen. Virol. (1992) 73:2785-2793
#title The complete nucleotide sequence of turnip mosaic potyvirus
  RNA.
#accession JQ1895
##molecule_type mRNA
##residues 1-3163 #label NTC
##cross-references DBJ:D10927; NID:g222660; PID:d1002200; PID:g222661
CLASSIFICATION #superfamily tobacco etch virus genome polyprotein; DEAB/H
  box helicase homology
KEYWORDS AIP; coat protein; genome-linked protein; inclusion protein;
  nucleus; P-loop; phosphoprotein; polyprotein
FEATURE
  1-362 #product P1 protein #status predicted #label PIP\
  363-820 #product helper component protein #status predicted
    #label HCP\
  821-1175 #product P3 protein #status predicted #label P3P\
  1176-1227 #product p6K1 protein #status predicted #label P6P\
  1228-1870 #product cytoplasmic inclusion protein #status predicted
    #label CIP\
  1313-1586 #domain DEAD/H box helicase homology #label DEAD\
  1313-1320 #region nucleotide-binding motif A (P-loop)\
  1398-1403 #region nucleotide-binding motif B\
  1402-1405 #region DEXH motif\
  1871-1923 #product p6K2 protein #status predicted #label PKP\
  1924-2115 #product VPg protein #status predicted #label VPG\

```

```

2115-2358      #product nuclear inclusion a protein #status predicted
               #label NIA\
2359-2875      #product nuclear inclusion b protein #status predicted
               #label NIB\
2876-3163      #product coat protein #status predicted #label CAP\
1986           #binding site phosphoryl-RNA (Tyr) (covalent) #status
               predicted
SUMMARY        #length 3163 #molecular-weight 357819 #checksum 4441

Query Match    62.2%; Score 51; DB 1; Length 3163;
Best Local Similarity 41.7%; Pred. No. 4.61e+00;
Matches 5; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

Db 1299 GKTAHSDKIL 1310
QY 1 GQGNPNRDIL 12

RESULT 12
ENTRY
TITLE      #type complete
ORGANISM   aspartate 1-decarboxylase pand - Bacillus subtilis
DATE       #formal_name Bacillus subtilis
           05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
           24-Sep-1998
ACCESSIONS A69672
REFERENCE   A69580
#authors    Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
            Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
            Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
            A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
            Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
            Chol, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
            Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.;
            Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.;
            Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
            M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Galleron, N.; Ghim,
            S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
            Guiseppi, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood,
            C.R.; Kenaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
            Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
            Kasahara, Y.; Klaerig-Blanchard, M.; Klein, C.; Kobayashi,
            Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.;
            Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
            Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
            Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
            M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
            M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro,
            V.; Pohl, T.M.; Portetelle, D.; Porwollik, S.; Prescott,
            A.M.; Pressecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
            Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
            Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.;
            Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
            Sekowska, A.; Seror, S.J.; Seror, P.; Shin, B.S.; Soldo,
            B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
            Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
            Terpestra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
            Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
            Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.;
            Winters, P.; Wipat, A.; Yamamoto, H.; Yanane, K.; Yasumoto,
            K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
            Yoshikawa, H.; Dauchin, A.
            Nature (1997) 390:249-256
#journal     The complete genome sequence of the Gram-positive bacterium
#title       Bacillus subtilis.
#cross-references M1UD:98044033
#accession   A69672
#status      preliminary; nucleic acid sequence not shown;
            translation not shown
#molecule_type DNA
#residues    1-127 #label KUN
#cross-references GB:Z99115; GB:AL009126; NID:c2634478; PID:e1183686;
            PID:g2634659
#experimental_source strain 168

```

```

GENETICS
#gene
CLASSIFICATION #superfamily aspartate 1-decarboxylase
SUMMARY        #length 127 #molecular-weight 13900 #checksum 7638

Query Match    61.0%; Score 50; DB 2; Length 127;
Best Local Similarity 60.0%; Pred. No. 7.43e+00;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 118 LGNEPARTIL 127
QY 3 IGNDPNRDIL 12

RESULT 13
ENTRY
TITLE      #type complete
ORGANISM   GTP cyclohydrolase I (EC 3.5.4.16) - Mycobacterium
           tuberculosis (strain H37RV)
DATE       #formal_name Mycobacterium tuberculosis
           17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
           21-Nov-1998
ACCESSIONS B70956
REFERENCE   A70500
#authors    Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
            C.; Harris, D.; Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry
            III, C.E.; Tekaiia, F.; Badcock, K.; Basham, D.; Brown, D.;
            Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
            Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
            Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;
            Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
            Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
            Skelton, S.; Squares, S.; Squires, R.; Sulston, J.E.;
            Taylor, K.; Whitehead, S.; Barrell, B.G.
            Nature (1998) 393:537-544
#journal     Deciphering the biology of Mycobacterium tuberculosis from
#title       the complete genome sequence.
#cross-references M1UD:98295987
#accession   B70956
#status      preliminary; nucleic acid sequence not shown;
            translation not shown
#molecule_type DNA
#residues    1-202 #label COL
#cross-references GB:Z95557; GB:AL123456; NID:g3242276; PID:e316840;
            PID:g2113967
#experimental_source strain H37RV

GENETICS
#gene
CLASSIFICATION #superfamily GTP cyclohydrolase I
KEYWORDS        isolate biosynthesis; GTP; hydrolase
SUMMARY        #length 202 #molecular-weight 22394 #checksum 3148

Query Match    61.0%; Score 50; DB 2; Length 202;
Best Local Similarity 70.0%; Pred. No. 7.43e+00;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 32 IGEDPPRDGL 41
QY 3 IGNDPNRDIL 12

RESULT 14
ENTRY
TITLE      #type complete
ALTERNATE_NAMES #formal_name Saccharomyces cerevisiae
                protein disulfide-isomerase (EC 5.3.4.1) precursor - Yeast
                (Saccharomyces cerevisiae)
                protein YCL043c; protein YCL1313; S-S rearrangase;
                thioredoxin-related glycoprotein 1; thyroid hormone-binding
                protein
ORGANISM      #formal_name Saccharomyces cerevisiae
DATE          30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change
            12-Jun-1998
ACCESSIONS    JX0182; P50224; A39376; S15050; JS0634; S40913; A41713;
            S19372; S25349
REFERENCE     JX0182

```

Tachikawa, H.; Miura, T.; Katakura, Y.; Mizunaga, T.
 J. Biochem. (1991) 110:306-313
 Molecular structure of a yeast gene, PDI1, encoding protein
 disulfide isomerase that is essential for cell growth.
 #cross-references MUID:92105067
 #accession JX0182
 #molecule_type DNA
 #residues 1-522 #label TAC
 #cross-references GB:D00842; NID:g218506; PID:d1001182; PID:g218507
 #accession PS0224
 #molecule_type protein
 #residues 99-112,'X',185-193;211-217,'X',233-237,'X',239;284-286,
 'X',288-298;309-315;325-335 #label TA2
 A39376
 LaMantia, M.; Miura, T.; Tachikawa, H.; Kaplan, H.A.;
 Lennarz, W.J.; Mizunaga, T.
 Proc. Natl. Acad. Sci. U.S.A. (1991) 88:4453-4457
 Glycosylation site binding protein and protein disulfide
 isomerase are identical and essential for cell viability in
 yeast.
 #cross-references MUID:91239586
 #accession A39376
 #molecule_type DNA
 #residues 1-522 #label LAM
 #cross-references GB:M62815
 S15050
 Scherens, B.; Dubois, E.; Messenguy, F.
 Yeast (1991) 7:185-193
 Determination of the sequence of the yeast YCL313 gene
 localized on chromosome III. Homology with the protein
 disulfide isomerase (PDI gene product) of other organisms.
 #cross-references MUID:91289690
 #accession S15050
 #molecule_type DNA
 #residues 1-522 #label SCH
 #cross-references EMBL:X57712; NID:g4801; PID:g4802
 JS0634
 Farquhar, R.; Honey, N.; Murant, S.J.; Bossier, P.; Schultz,
 L.; Montgomery, D.; Ellis, R.W.; Freedman, R.B.; Tuite,
 M.F.
 Gene (1991) 108:81-89
 Protein disulfide isomerase is essential for viability in
 Saccharomyces cerevisiae.
 #cross-references MUID:92104510
 #accession JS0634
 #molecule_type DNA
 #residues 1-113,'R',115-505,'EADAEAEA',506-522 #label FAR
 #cross-references EMBL:X54535; NID:g4119; PID:g4120
 S40913
 Kuentzel, H.
 Protein submitted to the EMBL Data Library, April 1990
 #accession S40913
 #molecule_type DNA
 #residues 1-82,'V',84-142,'S',144,146-167,'E',169-225,'V',227-457,
 'S',459-505,'EADAEAEA',506-522 #label KUE
 #cross-references EMBL:X52313; NID:g3948; PID:g3949
 A41713
 Guenther, R.; Brauer, C.; Janetzky, B.; Foerster, H.H.;
 Ebbrecht, I.M.; Lehle, L.; Kuentzel, H.
 J. Biochem. (1991) 286:24557-24563
 The Saccharomyces cerevisiae TRG1 gene is essential for
 growth and encodes a luminal endoplasmic reticulum
 glycoprotein involved in the maturation of vacuolar
 carboxypeptidase.
 #cross-references MUID:92105117
 #accession A41713
 #molecule_type DNA
 #residues 1-82,'V',84-142,'S',144,146-167,'E',169-225,'V',227-457,
 'S',459-505,'EADAEAEA',506-522 #label GUE
 #cross-references GB:M79982; NID:gl73023; PID:gl73024
 S19367
 Dubois, E.; Pierard, A.; Gigot, D.; Glansdorff, N.;
 Messenguy, F.; Scherens, B.

```
##cross-references EMBL:Z73517; NID:gl370341; PID:e247048; PID:gl370342;
MIPS:YPL161c
##experimental_source strain S288C (AB972)
REFERENCE S69428
#authors Purnelle, B.; Comblez, S.; Coster, F.; Naveau, F.; Goffeau,
A.
#submission submitted to the EMBL Data Library, March 1996
#description The sequence of 55 kb on the left arm of yeast chromosome XVI
identifies 28 open reading frames including 18 unknown
among which a new putative serine/threonine protein kinase,
a homologue to the human phosphotyrosyl phosphatase
activator PTPA and a homologue to the plant pleiotropic
regulator PRL1 of PPI and PP2a phosphatases.
#accession S69438
#molecule_type DNA
#residues 1-633 #label PUW
##cross-references EMBL:X96770; NID:gl403537; PID:e239041; PID:gl403548
GENETICS
#gene SGD:BEM4
#map_position 16L
#transmembrane protein
KEYWORDS
FEATURE 387-403
SUMMARY #length 633 #molecular_weight 70992 #status predicted #label TMM
#domain transmembrane
Query Match 61.0%; Score 50; DB 2; Length 633;
Best Local Similarity 70.0%; Pred. No. 7.43e+00;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Db 102 IGNDNREIL 111
| | | | |
Qy 3 IGNDNREIL 12
```

Search completed: Tue Aug 17 15:58:52 1999
Job time : 23 secs.

MPERCH_PP protein - protein database search, using Smith-Waterman algorithm

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Run on: Tue Aug 17 15:59:11 1999; MasPar time 2.83 Seconds
Tabular output not generated. 119.872 Million cell updates/sec

Title: >US-09-049-847-3
Description: (1-12) from US09049847.pep
Perfect Score: 82
Sequence: 1 GQIGNDPNRDL 12
Scoring table: PAM 150
Gap 15
Searched: 77977 seqs, 28268293 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries
Database: swiss-prot37
1:swissprot

Statistics: Mean 24.484; Variance 25.693; scale 0.953

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	82	100.0	1314	1	TETX_CLOTE	7.41e-09
2	52	63.4	403	1	PR36_DICDI	9.41e-01
3	52	63.4	413	1	PR36_SOLIU	9.41e-01
4	52	63.4	414	1	PR36_CAEEL	9.41e-01
5	52	63.4	418	1	PR36_RAT	9.41e-01
6	52	63.4	418	1	PR36_MOUSE	9.41e-01
7	52	63.4	418	1	PR36_HUMAN	9.41e-01
8	52	63.4	423	1	PR36_ASPNG	9.41e-01
9	52	63.4	428	1	PR36_YEAST	9.41e-01
10	52	63.4	3411	1	POLG_YEFV1	9.41e-01
11	52	63.4	3411	1	POLG_YEFV2	9.41e-01
12	51	62.2	190	1	GCH1_BACSU	1.60e+00
13	51	62.2	364	1	VMAT_NDV8	1.60e+00
14	51	62.2	375	1	GANA_PSEFL	1.60e+00
15	51	62.2	415	1	PR36_MANSE	1.60e+00
16	51	62.2	548	1	E13E_ARTSP	1.60e+00
17	51	62.2	548	1	E13E_ORFXA	1.60e+00
18	51	62.2	1115	1	PAN2_YEAST	1.60e+00
19	51	62.2	3163	1	POLG_TUMVQ	1.60e+00
20	51	62.2	3164	1	POLG_TUMVJ	1.60e+00
21	50	61.0	127	1	PAND_BACSU	2.71e+00
22	50	61.0	247	1	RHR_RHLIV	2.71e+00
23	50	61.0	522	1	PDI_YEAST	2.71e+00

24	50	61.0	633	1	BEM4_YEAST	BEM4 PROTEIN	2.71e+00
25	50	61.0	3770	1	ACVS_EMENI	DELTA-(L-ALPHA-AMINOAC	2.71e+00
26	49	59.8	841	1	CSW_DROME	PROTEIN-TYROSINE PHOSP	4.56e+00
27	48	58.5	179	1	GCHI_CAMGE	GTP CYCLOHYDROLASE I (7.58e+00
28	48	58.5	261	1	YEV6_YEAST	HYPOTHETICAL 29.7 KD P	7.58e+00
29	48	58.5	500	1	NU4C_TOBAC	NADH-PLASTOQUINONE OXI	7.58e+00
30	48	58.5	802	1	PUR2_YEAST	PHOSPHORIBOSYLAMINE-G	7.58e+00
31	48	58.5	1097	1	IMB3_HUMAN	IMPORIN BETA-3 SUBUNI	7.58e+00
32	47	57.3	84	1	UNGL_BPPB2	URACIL-DNA GLYCOSYLASE	1.25e+01
33	47	57.3	378	1	KLAE_ECOLI	KLAA PROTEIN (TELA PRO	1.25e+01
34	47	57.3	514	1	TFEB_HUMAN	TFEB PROTEIN (FRAGMENT	1.25e+01
35	47	57.3	796	1	DHG_ECOLI	GLUCOSE DEHYDROGENASE	1.25e+01
36	47	57.3	821	1	SPK1_YEAST	PROTEIN KINASE SPK1 (E	1.25e+01
37	47	57.3	900	1	KDPD_CLOAB	SENSOR PROTEIN KDPE (E	1.25e+01
38	47	57.3	1093	1	PIKA_DICDI	PHOSPHATIDYLINOSITOL 4	1.25e+01
39	47	57.3	1379	1	MET_MOUSE	HEPATOCYTE GROWTH FACT	1.25e+01
40	47	57.3	1772	1	MSPL_PLAYO	MEROPOITE SURFACE PROI	1.25e+01
41	46	56.1	397	1	TRPB_HAEIN	TRYPHOPHAN SYNTHASE BE	2.04e-01
42	46	56.1	417	1	CBPB_HUMAN	CARBOXYPEPTIDASE B PRE	2.04e-01
43	46	56.1	655	1	CDGT_KLEPN	CYCLOMALTODEXTRIN GLUC	2.04e-01
44	46	56.1	867	1	EF2_BLAHO	ELONGATION FACTOR 2 (E	2.04e+01
45	46	56.1	1184	1	BIMC_EMENI	KINESIN-LIKE PROTEIN B	2.04e+01

ALIGNMENTS

RESULT 1	TETX_CLOTE	STANDARD;	PRT; 1314 AA.
AC	P04958;		
DT	13-AUG-1987 (REL. 05, CREATED)		
DT	13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)		
DT	15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)		
DE	TETANUS TOXIN PRECURSOR (EC 3.4.24.68) (TENTOXILYSIN).		
OS	CLOSTRIDIUM TETANI.		
OG	PLASMID.		
OC	BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;		
OC	CLOSTRIDIUM.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE; 87053814.		
RA	EISEL U., JARAUSCH W., GORETZKI K., HENSCHEN A., ENGELS J.,		
RA	WELLER U., HUDEL M., HABERMANN E., NIEMANN H.,		
RT	"Tetanus toxin: primary structure, expression in E. coli, and		
RT	homology with botulinum toxins."		
RL	EMBO J. 5:2495-2502(1986).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-CN3911.		
RX	MEDLINE; 87040747.		
RA	FAIRWEATHER N.F., LYNES V.A.;		
RT	"The complete nucleotide sequence of tetanus toxin."		
RL	NUCLEIC ACIDS RES. 14:7809-7812(1986).		
RN	[3]		
RP	SEQUENCE OF 742-1314 FROM N.A.		
RX	MEDLINE; 86085672.		
RA	FAIRWEATHER N.F., LYNES V.A., PICKARD D.J., ALLEN G., THOMSON R.O.;		
RT	"Cloning, nucleotide sequencing, and expression of tetanus toxin		
RT	fragment C in Escherichia coli."		
RL	J. BACTERIOL. 165:21-27(1986).		
RN	[4]		
RP	PARTIAL SEQUENCE, AND DISULFIDE BONDS.		
RX	MEDLINE; 90201034.		
RA	KRIEGSTEIN K., HENSCHEN A., WELLER U., HABERMANN E.;		
RT	"Arrangement of disulfide bridges and positions of sulphydryl groups		
RT	in tetanus toxin."		
RL	EUR. J. BIOCHEM. 188:39-45(1990).		
RN	[5]		
RP	PARTIAL SEQUENCE.		
RX	MEDLINE; 92037649.		
RA	KRIEGSTEIN K.G., HENSCHEN A.H., WELLER U., HABERMANN E.;		
RT	"Limited proteolysis of tetanus toxin. Relation to activity and		
RT	identification of cleavage sites."		
RL	EUR. J. BIOCHEM. 202:41-51(1991).		

RN IDENTIFICATION AS ZINC-PROTEASE.
 RX MEDLINE; 93010948.
 RA SCHIAVO G., POULAIN B., ROSSETTO O., BENFENATI F., TAUC L.,
 RA MONTECUCCO C.;
 RI "Tetanus toxin is a zinc protein and its inhibition of
 RI neurotransmitter release and protease activity depend on zinc.";
 RL EMBO J. 11:3577-3583(1992).
 [6]
 RN IDENTIFICATION OF SUBSTRATE.
 RX MEDLINE; 93053295.
 RA SCHIAVO G., BENFENATI F., POULAIN B., ROSSETTO O., DE LAURETO P.P.,
 RA DASGUPTA B.R., MONTECUCCO C.;
 RI "Tetanus and botulinum-B neurotoxins block neurotransmitter release
 RI by proteolytic cleavage of synaptobrevin.";
 RL NATURE 359:832-835(1992).
 [8]
 RN X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 874-1314.
 RX MEDLINE; 97475217.
 RA UMLAND T.C., WINGERT L.M., SWAMINATHAN S., FUREY W.F., SCHMIDT J.J.,
 RA SAX M.;
 RI "Structure of the receptor binding fragment HC of tetanus
 RI neurotoxin.";
 RL NAT. STRUCT. BIOL. 4:788-792(1997).
 CC -!- FUNCTION: TETANUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
 CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
 CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
 CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
 CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
 CC ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 76-GLN-|-PHE-77
 CC BOND OF SYNAPTOSOMAL-2.
 CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF 76-GLN-|-PHE-77 BOND IN
 CC SYNAPTOSOMAL-2.
 CC -!- SUBUNIT: THE PRECURSOR POLYPEPTIDE IS SUBSEQUENTLY CLEAVED TO
 CC YIELD SUBCHAINS L AND H. THESE REMAIN LINKED BY A DISULFIDE BRIDGE
 CC AND ARE NON-TOXIC AFTER SEPARATION.
 CC -!- THE C-TERMINAL OF THE HEAVY CHAIN BINDS TO GANGLIOSIDE RECEPTORS.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27 (ZINC
 CC METALLOPROTEASES); ALSO KNOWN AS THE TETANUS/BOTULINUM NEUROTOXIN
 CC SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; X04436; G40770; -.
 DR EMBL; M12739; G144921; -.
 DR EMBL; X06214; G40774; -.
 DR PIR; A25689; BTCLIN.
 DR PDB; 1AF9; 29-APR-98.
 DR PROSITE; PS00142; ZINC_PROTEASE; 1.
 KW NEUROTOXIN; TRANSMEMBRANE; HYDROLASE; METALLOPROTEASE; ZINC; PLASMID;
 KW 3D-STRUCTURE.
 FT INIT_MET 0 0
 FT CHAIN 1 456
 FT CHAIN 457 1314
 FT METAL 232 232
 FT ACT_SITE 233 233
 FT METAL 236 236
 FT TRANSMEM 226 246
 FT TRANSMEM 689 689
 FT DISULFID 438 466
 FT DISULFID 1076 1092
 SQ SEQUENCE 1314 AA; 150550 MW; 26190E3E CRC32;

Query Match 100.0%; Score 82; DB 1; Length 1314;
 Best Local Similarity 100.0%; Pred. No. 7.4le-09;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1272 GQIGNDPNRDIL 1283
 |||||
 QY 1 GQIGNDPNRDIL 12
 RESULT 2
 ID PRS6_DICDI STANDARD; PRT; 403 AA.
 AC P34123;
 DT 01-FEB-1994 (REL. 28, CREATED)
 DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE 26S PROTEASE REGULATORY SUBUNIT 6B HOMOLOG (TAT-BINDING PROTEIN
 DE HOMOLOG 2).
 DE TBPB OR TBP2.
 GN DICTYOSTELIUM DISCOIDEUM (SLIME MOLD).
 OS EUKARYOTA; DICTYOSTELIIDA; DICTYOSTELIUM.
 OC
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AX4;
 RX MEDLINE; 93312322.
 RA SHAW D.R., ENNIS H.L.;
 RI "Molecular cloning and developmental regulation of Dictyostelium
 RI discoideum homologues of the human and yeast HIV1 Tat-binding
 RI protein.";
 RL BIOCHEM. BIOPHYS. RES. COMMUN. 193:1291-1296(1993).
 CC -!- FUNCTION: THE 26S PROTEASE IS INVOLVED IN THE ATP-DEPENDENT
 CC DEGRADATION OF UBQUITINATED PROTEINS. THE REGULATORY (OR ATPASE)
 CC COMPLEX CONFERS ATP DEPENDENCY AND SUBSTRATE SPECIFICITY TO THE
 CC 26S COMPLEX (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR (POTENTIAL).
 CC -!- DEVELOPMENTAL STAGE: HIGHEST EXPRESSION IN VEGETATIVELY GROWING
 CC CELLS. THE LEVEL OF EXPRESSION FALLS STEADILY THROUGHOUT
 CC MULTICELLULAR DEVELOPMENT AND ARE NOT FOUND IN DORMANT OR
 CC GERMINATING SPORES.
 CC -!- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; L16578; G290055; -.
 DR PIR; JN0611; JN0611.
 DR DICTYDB; DD01052; TBPB.
 DR PROSITE; PS00674; AAA; 1.
 DR PFAM; PF00004; AAA; 1.
 KW PROTEASOME; ATP-BINDING; NUCLEAR PROTEIN.
 FT NP_BIND 191 198
 FT ATP (POTENTIAL).
 SQ SEQUENCE 403 AA; 45542 MW; F1A7C9A7 CRC32;
 Query Match 63.4%; Score 52; DB 1; Length 403;
 Best Local Similarity 63.6%; Pred. No. 9.4le-01;
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 178 QIGIDPPRGVL 188
 |||||
 QY 2 QIGIDPPNRDIL 12
 RESULT 3
 ID PRS5_SOLTU STANDARD; PRT; 413 AA.
 AC P54778;
 DT 01-OCT-1996 (REL. 34, CREATED)
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE 26S PROTEASE REGULATORY SUBUNIT 6B HOMOLOG.
 OS SOLANUM TUBEROSUM (POTATO).
 OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
 OC EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;
 OC ASTERIDAE; SOLANACEAE; SOLANALES; SOLANACEAE; SOLANUM.

```

RN  SEQUENCE FROM N.A.
RP  STRAIN=CV. SUPERIOR;
RA  HART J.K., HANNAPEL D.J.;
RT  "Isolation of an AAA superfamily cDNA clone from potato.";
RL  (IN) PLANT GENE REGISTER PGR96-039.
CC  -!- FUNCTION: THE 26S PROTEASE IS INVOLVED IN THE ATP-DEPENDENT
CC  DEGRADATION OF UBQUITINATED PROTEINS. THE REGULATORY (OR ATPASE)
CC  COMPLEX CONFERS ATP DEPENDENCY AND SUBSTRATE SPECIFICITY TO THE
CC  26S COMPLEX (BY SIMILARITY).
CC  -!- SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR (POTENTIAL).
CC  -!- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; U43398; G115334; -.
DR  PROSITE; PS00674; AAA; 1.
DR  PFAM; PF00004; AAA; 1.
KW  PROTEASOME; ATP-BINDING; NUCLEAR PROTEIN.
FT  NP_BIND 201 208 ATP (POTENTIAL).
SQ  SEQUENCE 413 AA; 46532 MW; C75308E2 CRC32;

Query Match 63.4%; Score 52; DB 1; Length 413;
Best Local Similarity 63.6%; Pred. No. 9.41e-01;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 188 QIGDPPRGVL 198
   ||| ||| |
QY 2 QIGNDPNRDL 12

RESULT 4
ID PRS6_CABEL STANDARD; PRT; 414 AA.
AC P46502;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE PROBABLE 26S PROTEASE REGULATORY SUBUNIT 6B.
GN F23F12.6.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RA DU Z.;
RC STRAIN=BRISTOL N2;
RL SUBMITTED (AUG-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- FUNCTION: THE 26S PROTEASE IS INVOLVED IN THE ATP-DEPENDENT
CC DEGRADATION OF UBQUITINATED PROTEINS. THE REGULATORY (OR ATPASE)
CC COMPLEX CONFERS ATP DEPENDENCY AND SUBSTRATE SPECIFICITY TO THE
CC 26S COMPLEX (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR  EMBL; U12965; G529215; -.
DR  WORMPEP; F23F12.6; C601253.
DR  PROSITE; PS00674; AAA; 1.
DR  PFAM; PF00004; AAA; 1.
KW  HYPOTHETICAL PROTEIN; PROTEASOME; ATP-BINDING; NUCLEAR PROTEIN.

```

```

FT  NP_BIND 202 209 ATP (POTENTIAL).
SQ  SEQUENCE 414 AA; 46538 MW; 51F26490 CRC32;

Query Match 63.4%; Score 52; DB 1; Length 414;
Best Local Similarity 63.6%; Pred. No. 9.41e-01;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 189 QIGDPPRGVL 199
   ||| ||| |
QY 2 QIGNDPNRDL 12

RESULT 5
ID PRS6_RAT STANDARD; PRT; 418 AA.
AC Q63570;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE 26S PROTEASE REGULATORY SUBUNIT 6B (IAT-BINDING PROTEIN-7) (TBP-7).
GN PSMC4.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=SPRAGUE-DAWLEY; TISSUE=LIVER;
RX MEDLINE; 96183075.
RA MAKINO I., YOGOSAWA S., KANEMAKI M., YOSHIDA T., YAMANO K.,
RA KISHIMOTO T., MONCOLLIN V., EGY J.M., MURAMATSU M., TAMURA T.;
RT "Structures of the rat proteasomal ATPases: determination of highly
RT conserved structural motifs and rules for their spacing.";
RL BIOCHEM. BIOPHYS. RES. COMMUN. 220:1049-1054(1996).
CC -!- FUNCTION: THE 26S PROTEASE IS INVOLVED IN THE ATP-DEPENDENT
CC DEGRADATION OF UBQUITINATED PROTEINS. THE REGULATORY (OR ATPASE)
CC COMPLEX CONFERS ATP DEPENDENCY AND SUBSTRATE SPECIFICITY TO THE
CC 26S COMPLEX.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR.
CC -!- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR  EMBL; D50695; G1395186; -.
DR  PROSITE; PS00674; AAA; 1.
DR  PFAM; PF00004; AAA; 1.
KW  PROTEASOME; ATP-BINDING; NUCLEAR PROTEIN.
FT  NP_BIND 206 213 ATP (POTENTIAL).
SQ  SEQUENCE 418 AA; 47408 MW; 06A3A60B CRC32;

Query Match 63.4%; Score 52; DB 1; Length 418;
Best Local Similarity 63.6%; Pred. No. 9.41e-01;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 193 QIGDPPRGVL 203
   ||| ||| |
QY 2 QIGNDPNRDL 12

RESULT 6
ID PRS6_MOUSE STANDARD; PRT; 418 AA.
AC P54775;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE 26S PROTEASE REGULATORY SUBUNIT 6B (CIP21).
GN PSMC4.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;

```

CC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RA CHOI H.S., SEOL W., MOORE D.D.;
RL SUBMITTED (FEB-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- FUNCTION: THE 26S PROTEASE IS INVOLVED IN THE ATP-DEPENDENT
CC DEGRADATION OF UBQUITINATED PROTEINS. THE REGULATORY (OR ATPASE)
CC COMPLEX CONFERS ATP DEPENDENCY AND SUBSTRATE SPECIFICITY TO THE
CC 26S COMPLEX. INTERACTS WITH THE ORPHAN NUCLEAR HORMONE RECEPTOR
CC MB67.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR.
CC -!- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
CC or send an email to license@isb-sib.ch.
CC
CC EMBL; L76223; G1196528; -.
DR PROSITE; PS00674; AAA; 1.
DR PFAM; PF00004; AAA; 1.
KW PROTEASOME; ATP-BINDING; NUCLEAR PROTEIN.
FT NP_BIND 206 213 ATP (POTENTIAL).
SQ SEQUENCE 418 AA; 47281 MW; 8DCA5892 CRC32;

Query Match 53.4%; Score 52; DB 1; Length 418;
Best Local Similarity 53.6%; Pred. No. 9.41e-01;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 193 QIGIDPPRGVL 203
QY 2 QIGNDPNRDIL 12
|||||

RESULT 7
ID PRS6_HUMAN STANDARD; PRT; 418 AA.
AC P43686;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE 26S PROTEASE REGULATORY SUBUNIT 6B (TAT-BINDING PROTEIN-7) (TBP-7).
GN PSMC4 OR TBP7.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93126329.
RA OHANA B., MOORE P.A., RUBEN S.M., SOUTHGATE C.D., GREEN M.R.,
RA ROSEN C.A.;
RT "The type I human immunodeficiency virus Tat binding protein is a
RT transcriptional activator belonging to an additional family of
RT evolutionarily conserved genes."
RL PROC. NATL. ACAD. SCI. U.S.A. 90:138-142(1993).
RN [2]
RP PARTIAL SEQUENCE, REVISION TO C-TERMINAL, AND FUNCTION.
RX MEDLINE; 9438582.
RA DUBIEL W., FERRELL K., RECHSTEINER M.;
RT "Tat-binding protein 7 is a subunit of the 26S protease.";
RL BIOL. CHEM. HOPPE-SEYLER 375:237-240(1994).
CC -!- FUNCTION: THE 26S PROTEASE IS INVOLVED IN THE ATP-DEPENDENT
CC DEGRADATION OF UBQUITINATED PROTEINS. THE REGULATORY (OR ATPASE)
CC COMPLEX CONFERS ATP DEPENDENCY AND SUBSTRATE SPECIFICITY TO THE
CC 26S COMPLEX.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR.
CC -!- PTM: THE N-TERMINAL IS BLOCKED.
CC -!- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
DR MM; 602707; -.
DR PROSITE; PS00674; AAA; 1.

DR PFAM; PF00004; AAA; 1.
KW PROTEASOME; ATP-BINDING; NUCLEAR PROTEIN.
FT NP_BIND 206 213 ATP (POTENTIAL).
SQ SEQUENCE 418 AA; 47336 MW; 4D06C2AF CRC32;

Query Match 63.4%; Score 52; DB 1; Length 418;
Best Local Similarity 63.6%; Pred. No. 9.41e-01;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 193 QIGIDPPRGVL 203
QY 2 QIGNDPNRDIL 12
|||||

RESULT 8
ID PRS6_ASPNG STANDARD; PRT; 423 AA.
AC P78578;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE 26S PROTEASE REGULATORY SUBUNIT 6B HOMOLOG.
GN TBP4.
OS ASPERGILLUS NIGER.
OC EUKARYOTA; FUNGI; ASCOMYCOTA; EUASCOMYCETES; PLECTOMYCETES;
OC EUKARYOTA; FUNGI; TRICHOCOMACEAE; MITOSPORIC TRICHOCOMACEAE; ASPERGILLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CBS 120.49 / N400;
RA DELRU C., JARAI G., BUXTON F.;
RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- FUNCTION: THE 26S PROTEASE IS INVOLVED IN THE ATP-DEPENDENT
CC DEGRADATION OF UBQUITINATED PROTEINS. THE REGULATORY (OR ATPASE)
CC COMPLEX CONFERS ATP DEPENDENCY AND SUBSTRATE SPECIFICITY TO THE
CC 26S COMPLEX (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
CC or send an email to license@isb-sib.ch.
CC
CC EMBL; U15601; G1777414; -.
DR PROSITE; PS00674; AAA; 1.
DR PFAM; PF00004; AAA; 1.
KW PROTEASOME; ATP-BINDING; NUCLEAR PROTEIN.
FT NP_BIND 207 214 ATP (POTENTIAL).
SQ SEQUENCE 423 AA; 47222 MW; 14E49380 CRC32;

Query Match 63.4%; Score 52; DB 1; Length 423;
Best Local Similarity 63.6%; Pred. No. 9.41e-01;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 194 QIGIDPPRGVL 204
QY 2 QIGNDPNRDIL 12
|||||

RESULT 9
ID PRS6_YEAST STANDARD; PRT; 428 AA.
AC P33298;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE 26S PROTEASE REGULATORY SUBUNIT 6B HOMOLOG (INT1 PROTEIN) (TAT-BINDING
DE HOMOLOG 2).
GN YTA2 OR INT1 OR YDR394W OR D9509.14.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.

[1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=C836;
 RX MEDLINE; 95274317.
 RA SCHNALL R., MANNHAUPT G., STUCKA R., TAUER R., EHMLE S.,
 RA SCHWARZLOSE C., VETTER I., FELDWMANN H.;
 RT "Identification of a set of yeast genes coding for a novel family of
 RT putative ATPases with high similarity to constituents of the 26S
 RT protease complex."
 RL YEAST 10:1141-1155(1994).
 RN [2]
 RN SEQUENCE FROM N.A.
 RP CAMPBELL C.L., TANAKA N., WHITE K.H., THORSNESS P.E.;
 RL SUBMITTED (FEB-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [3]
 RN SEQUENCE FROM N.A.
 RA DIETRICH F.S., MULLIGAN J., ALLEN E., ARAUJO R., AVILES E.,
 RA BERNO A., CARPENTER J., CHEN E., CHERRY J.M., CHUNG E., DUNCAN M.,
 RA HUNICKE-SMITH S., HYMAN R., KOMP C., LASHKARI D., LEW H., LIN D.,
 RA MOSEDALE D., NAKAHARA K., NAWATH A., OEFNER P., OH C., PETEL F.X.,
 RA ROBERTS D., SCHRAMM S., SCHROEDER M., SHOGREN T., SHROFF N.,
 RA WINANT A., YELTON M., BOISTEIN D., DAVIS R.W.;
 RL SUBMITTED (JUL-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC -1- FUNCTION: THE 26S PROTEASE IS INVOLVED IN THE ATP-DEPENDENT
 CC DEGRADATION OF UBQUITINATED PROTEINS. THE REGULATORY (OR ATPASE)
 CC COMPLEX CONFERS ATP DEPENDENCY AND SUBSTRATE SPECIFICITY TO THE
 CC 26S COMPLEX (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X73570; G313880; -.
 DR EMBL; U06229; G458389; -.
 DR EMBL; U32274; G927327; -.
 DR PIR; S45606; S46606;
 DR SGD; L0002556; YTA2.
 DR PROSITE; PS00674; AAA; 1.
 DR PFAM; PF00004; AAA; 1.
 KW PROTEASOME; ATP-BINDING; NUCLEAR PROTEIN.
 FT NP_BIND 213 220 ATP (POTENTIAL).
 FT CONFLICT 342 342 Y->S (IN REF. 3).
 SQ SEQUENCE 428 AA; 47969 MW; E96A890E CRC32;
 Query Match 63.4%; Score 52; DB 1; Length 428;
 Best Local Similarity 63.6%; Pred. No. 9.41e-01;
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 Db 200 QIGIDPRGV 210
 QY 2 QIGNDPNRDL 12
 RESULT 10
 ID POLG_YEYV1 STANDARD; PRT; 3411 AA.
 AC P03314; O42028;
 DT 21-JUL-1986 (REL. 01, CREATED)
 DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
 DE GENOME POLYPEPTIDE [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN); MATRIX
 DE PROTEIN (ENVELOPE PROTEIN M); MAJOR ENVELOPE PROTEIN E; NONSTRUCTURAL
 DE PROTEINS NS1, NS2A, NS2B, NS4A AND NS4B; HELICASE (NS3); RNA-DIRECTED
 DE RNA POLYMERASE (EC 2.7.7.48) (NS5)].
 OS YELLOW FEVER VIRUS (STRAIN 17D).
 OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; FLAVIVIRIDAE;
 OC FLAVIVIRUS.
 RN [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE; 85272570.
 RA RICE C.M., LENCHE E.M., EDDY S.R., SHIN S.J., SHEETS R.L.,
 RA STRAUSS J.H.;
 RT "Nucleotide sequence of yellow fever virus: implications for
 RT flavivirus gene expression and evolution."
 RL SCIENCE 229:726-733(1985).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X03700; G59339; -.
 DR EMBL; K02749; G336193; -.
 DR PIR; A03914; GNMVY.
 DR PFAM; PF00869; Flavi_glycoprot; 1.
 DR PFAM; PF00948; Flavi_NS1; 1.
 DR PFAM; PF00949; Flavi_helicase; 1.
 DR PFAM; PF00972; Flavi_NS5; 1.
 DR PFAM; PF01002; Flavi_NS2B; 1.
 DR PFAM; PF01003; Flavi_capsid; 1.
 DR PFAM; PF01004; Flavi_M; 1.
 DR PFAM; PF01005; Flavi_NS2A; 1.
 DR HSP; P14336; 1SVB.
 KW POLYPEPTIDE; GLYCOPROTEIN; TRANSFERASE; RNA-DIRECTED RNA POLYMERASE;
 KW CORE PROTEIN; COAT PROTEIN; ENVELOPE PROTEIN; HELICASE; ATP-BINDING;
 KW TRANSMEMBRANE; NONSTRUCTURAL PROTEIN.
 FT INIT_MET 1 1
 FT CHAIN 1 121
 FT PROPEP 122 210
 FT CHAIN 211 285
 FT CHAIN 286 778
 FT CHAIN 779 1130
 FT CHAIN 1131 1354
 FT CHAIN 1355 1484
 FT CHAIN 1485 2107
 FT CHAIN 2108 2256
 FT CHAIN 2257 2506
 FT CHAIN 2507 3411
 FT TRANSMEM 249 269
 FT TRANSMEM 271 285
 FT TRANSMEM 740 753
 FT TRANSMEM 755 778
 FT TRANSMEM 1159 1180
 FT DOMAIN 383 396
 FT NP_BIND 1682 1689
 FT SITE 1773 1776
 FT DISULFID 248 315
 FT DISULFID 345 401
 FT DISULFID 359 390
 FT DISULFID 377 406
 FT DISULFID 467 568
 FT DISULFID 585 615
 FT CARBOHYD 134 134
 FT CARBOHYD 150 150
 FT CARBOHYD 908 908
 FT CARBOHYD 986 986
 FT CARBOHYD 2320 2320
 FT CARBOHYD 2346 2346
 FT CARBOHYD 2467 2467
 SQ SEQUENCE 3411 AA; 379512 MW; E85D316D CRC32;
 CAPSID PROTEIN C.
 ENVELOPE GLYCOPROTEIN M.
 MAJOR ENVELOPE PROTEIN E.
 NONSTRUCTURAL PROTEIN NS1.
 NONSTRUCTURAL PROTEIN NS2A.
 NONSTRUCTURAL PROTEIN NS2B.
 HELICASE (NS3).
 NONSTRUCTURAL PROTEIN NS4A.
 NONSTRUCTURAL PROTEIN NS4B.
 RNA-DIRECTED RNA POLYMERASE (NS5).
 POTENTIAL.
 POTENTIAL.
 POTENTIAL.
 POTENTIAL.
 POTENTIAL.
 INVOLVED IN FUSION.
 ATP (POTENTIAL).
 DEAH BOX.
 BY SIMILARITY.
 BY SIMILARITY.
 BY SIMILARITY.
 BY SIMILARITY.
 BY SIMILARITY.
 POTENTIAL.
 POTENTIAL.
 POTENTIAL.
 POTENTIAL.
 POTENTIAL.
 POTENTIAL.

```

FT NP_BIND 1689 ATP (POTENTIAL).
FT SITE 1773 DEAH_BOX.
FT TRANSMEM 249 POTENTIAL.
FT TRANSMEM 271 POTENTIAL.
FT TRANSMEM 740 POTENTIAL.
FT TRANSMEM 755 POTENTIAL.
FT TRANSMEM 1159 POTENTIAL.
FT DISULFID 288 BY SIMILARITY.
FT DISULFID 345 BY SIMILARITY.
FT DISULFID 359 BY SIMILARITY.
FT DISULFID 377 BY SIMILARITY.
FT DISULFID 467 BY SIMILARITY.
FT DISULFID 585 BY SIMILARITY.
FT CARBOHYD 134 POTENTIAL.
FT CARBOHYD 150 POTENTIAL.
FT CARBOHYD 908 POTENTIAL.
FT CARBOHYD 986 POTENTIAL.
FT CARBOHYD 2320 POTENTIAL.
FT CARBOHYD 2346 POTENTIAL.
FT CARBOHYD 2467 POTENTIAL.
FT SEQUENCE 3411 AA; 379524 MW; ADCCF22B CRC32;

Query Match 53.4%; Score 52; DB 1; Length 3411;
Best Local Similarity 70.0%; Pred. No. 9.41e-01;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0

Db 1947 GRIGNPNRD 1956
QY 1 GQIGNPNRD 10
      !:!!:!!!!

RESULT 12
ID GCH1_BACSU STANDARD; PRT; 190 AA.
AC P19455;
DT 01-FEB-1991 (REL. 17, CREATED)
DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE GTP CYCLOHYDROLASE I (EC 3.5.4.16) (GTP-CH-I).
DE MTRA.
GN BACILLUS SUBTILIS.
OS BACILLUS SUBTILIS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE; 91062353.
RX GOLLNICK P., ISHINO S., KURODA M.I., HENNER D.J., YANOFSKY C.;
RA "The mtr locus is a two-gene operon required for transcription
RT attenuation in the trp operon of Bacillus subtilis.";
RL PROC. NAFL. ACAD. SCI. U.S.A. 87:8726-8730(1990).
[2]
RN SEQUENCE OF 1-19 FROM N.A.
RP STRAIN=168 / JH642;
RC MEDLINE; 91216992.
RA MICKA B., GROCH N., HEINEMANN U., MARAHIEL M.A.;
RT "Molecular cloning, nucleotide sequence, and characterization of the
RL Bacillus subtilis gene encoding the DNA-binding protein HBSu.";
J. BACTERIOL. 173:3191-3198(1991).
[3]
RN FUNCTION.
RP MEDLINE; 92202128.
RX BABITZKE P., GOLLNICK P., YANOFSKY C.;
RA "The mtrAB operon of Bacillus subtilis encodes GTP cyclohydrolase I
RT (MtrA), an enzyme involved in folic acid biosynthesis, and MtrB, a
RT regulator of tryptophan biosynthesis.";
RL J. BACTERIOL. 174:2059-2064(1992).
[4]
RN CHARACTERIZATION.
RP MEDLINE; 95194311.
RX DE SAIZUE A., VANKAN P., VAN LOON A.P.;
RA "Enzymic characterization of Bacillus subtilis GTP cyclohydrolase I.
RT Evidence for a chemical dephosphorylation of dihydroneopterin
RT triphosphate.";
RL BIOCHEM. J. 306:371-377(1995).

```

```

CC      -!- CATALYTIC ACTIVITY: GTP + 2 H(2)O -> FORMATE + 2-AMINO-4-HYDROXY-
CC      6-(BRYTHRO-1,2,3-TRIHYDROXYPROPYL)DIHYDROPTERIDINE TRIPHOSPHATE.
CC      -!- ENZYME REGULATION: K+ IONS MODERATELY INCREASES THE VMAX., WHEREAS
CC      UTP AND CA2+ AND MG2+ IONS DRASTICALLY INCREASE THE KM FOR GTP.
CC      -!- PATHWAY: FIRST STEP IN THE BIOSYNTHESIS OF TETRAHYDROFOLATE.
CC      -!- SUBUNIT: HOMOPOLYMER.
CC      -!- SIMILARITY: BELONGS TO THE GTP CYCLOHYDROLASE I FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; M37320; G143231; -
CC      EMBL; M80245; G143799; -
CC      EMBL; X52418; -; NOT_ANNOTATED_CDS.
CC      FIRM; A38256; A38256.
CC      FIRM; A39409; A39409.
CC      SUBILIST; BG10277; MTRA.
CC      PROSITE; PS00859; GTP_CYCLOHYDROL_1_1; 1.
CC      PROSITE; PS00860; GTP_CYCLOHYDROL_1_2; 1.
CC      PFAM; PF01227; GTP_CYCLOHYDROL; 1.
CC      ONE-CARBON METABOLISM; HYDROLASE; ALLOSTERIC ENZYME.
CC      DISULFID 78 150 BY SIMILARITY.
CC      SQ SEQUENCE 190 AA; 21219 MW; 2790F255 CRC32;
CC
CC      Query Match 62.2%; Score 51; DB 1; Length 190;
CC      Best Local Similarity 70.0%; Pred. No. 1.60e+00;
CC      Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
CC
CC      Db 20 IGDPNREGL 29
CC      QY 3 IGNDPNRDL 12
CC
CC      RESULT 13
CC      ID VVAT_NDVB STANDARD; PRT; 364 AA.
CC      AC P06157;
CC      DT 01-JAN-1988 (REL. 06, CREATED)
CC      DT 01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)
CC      DT 01-AUG-1992 (REL. 23, LAST ANNOTATION UPDATE)
CC      DE MATRIX PROTEIN.
CC      GN M.
CC      OS NEWCASTLE DISEASE VIRUS (STRAIN BEAUDETTE C/45) (NDV).
CC      OC VIRUSES; SSRNA NEGATIVE-STRAND VIRUSES; MONONEGAVIRALES;
CC      OC PARAMYXOVIRIDAE; PARAMYXOVIRINAE; RUBULAVIRUS.
CC      [1]
CC      RP SEQUENCE FROM N.A.
CC      RX MEDLINE; 87066775.
CC      RA CHAMBERS P., MILLAR N.S., PLATT S.G., EMMERSON P.T.;
CC      RT "Nucleotide sequence of the gene encoding the matrix protein of
CC      RT Newcastle disease virus."
CC      RL NUCLEIC ACIDS RES. 14:9051-9061(1986).
CC      -!- FUNCTION: THE M PROTEIN HAS A CRUCIAL ROLE IN VIRUS ASSEMBLY
CC      AND INTERACTS WITH THE RNP COMPLEX AS WELL AS WITH THE VIRAL
CC      MEMBRANE.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; X04687; G60941; -
CC      FIRM; A26111; MFNZNC.
CC      PFAM; PF00661; Matrix; 1.
CC      MATRIX PROTEIN; ENVELOPE PROTEIN.
CC      SQ SEQUENCE 364 AA; 39605 MW; 635391DD CRC32;

```

```

Query Match 62.2%; Score 51; DB 1; Length 364;
Best Local Similarity 41.7%; Pred. No. 1.60e+00;
Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 76 GMINDMPKRELL 87
QY 1 GQIGNDPNRDL 12

RESULT 14
ID GANA_PSEFL STANDARD; PRT; 376 AA.
AC P48841;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE ARABINOGALACTAN ENDO-1,4-BETA-GALACTOSIDASE PRECURSOR (EC 3.2.1.89)
DE (ENDO-1,4-BETA-GALACTANASE) (GALACTANASE).
GN GANA.
OS PSEUDOMONAS FLUORESCENS.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PSEUDOMONAS GROUP;
OC PSEUDOMONAS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SP. CELLULOSA;
RA BRAITHWAITE K.L., GILBERT H.J.;
RL SUBMITTED (SEP-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GALACTOSIDIC
CC LINKAGES IN ARABINOGALACTANS.
CC -!- SIMILARITY: BELONGS TO FAMILY 53 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X91885; G1017437; -
CC KW HYDROLASE; GLYCOSIDASE; SIGNAL.
CC FT SIGNAL 1 17 POTENTIAL.
CC FT CHAIN 18 376 ARABINOGALACTAN ENDO-1,4-BETA-
CC FT GALACTOSIDASE.
CC SQ SEQUENCE 376 AA; 42315 MW; 56D4C50F CRC32;
CC
CC      Query Match 62.2%; Score 51; DB 1; Length 376;
CC      Best Local Similarity 54.5%; Pred. No. 1.60e+00;
CC      Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
CC
CC      Db 157 OYGNENINIEL 167
CC      QY 2 QIGNDPNRDL 12
CC
CC      RESULT 15
CC      ID PRS6_MANSE STANDARD; PRT; 415 AA.
CC      AC P46507;
CC      DT 01-NOV-1995 (REL. 32, CREATED)
CC      DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
CC      DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
CC      DE 26S PROTEASE REGULATORY SUBUNIT 6S (ATPASE MS73).
CC      OS MANDUCA SEXTA (TOBACCO HAWMOTH) (TOBACCO HORNWORM).
CC      OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
CC      OC PTERYGOTA; LEPIDOPTERA; SPHINGIODEA; SPHINGINAE; MANDUCA.
CC      [1]
CC      RP SEQUENCE FROM N.A.
CC      RC TISSUE=INTERSEGMENTAL MUSCLE;
CC      RX MEDLINE; 95130567.
CC      RA DAWSON S.P., ARNOLD J., MAYER N.J., REYNOLDS S., BILLET M.A.,
CC      RA GORDON C., COLLEAUX L., KLOETZEL P., TANAKA K., MAYER R.J.;
CC      RT "Developmental changes of the 26 S proteasome in abdominal
CC      RT intersegmental muscles of Manduca sexta during programmed cell

```

```

RT death.";
RL J. BIOL. CHEM. 270:1850-1858(1995).
CC -!- FUNCTION: THE 26S PROTEASE IS INVOLVED IN THE ATP-DEPENDENT
CC DEGRADATION OF DEUTERATED PROTEINS. THE REGULATORY (OR ATPASE)
CC COMPLEX CONFERS ATP DEPENDENCY AND SUBSTRATE SPECIFICITY TO THE
CC 26S COMPLEX.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR (POTENTIAL).
CC -!- DEVELOPMENTAL STAGE: EXPRESSION OF THE GENE DRAMATICALLY INCREASES
CC IN THE PRE-ECLOSURE PERIOD.
CC -!- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z38135; G559486; -.
DR PROSITE; PS00674; AAA; 1.
DR PFAM; PF00004; AAA; 1.
KW PROTEASOME; ATP-BINDING; NUCLEAR PROTEIN.
FT NP_BIND 203 210 ATP (POTENTIAL).
SQ SEQUENCE 415 AA; 47053 MW; 7BD9FB35 CRC32;

Query Match      62.2%; Score 51; DB 1; Length 415;
Best Local Similarity 54.5%; Pred. No. 1.60e+00;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 190 QIGIEPPRGVL 200
QY 2 QIGNDPNRDIL 12
|||:|:|

```

Search completed: Tue Aug 17 15:59:18 1999
 Job time : 7 secs.

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Tue Aug 17 15:59:34 1999; MasPar time 5.83 Seconds
112.413 Million cell updates/sec
Tabular output not generated.

Title: >US-09-049-847-3
Description: (1-12) from US09049847.pep
Perfect Score: 82
Sequence: 1 GQIGNDPNRDIL 12

Scoring table: PAM 150
Gap 15

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: spiremb19

1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human
5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organella
9:sp-phage 10:sp-plant 11:sp-rodent 12:sp-unclassified
13:sp-vertebrate 14:sp-virus

Statistics: Mean 23.731; Variance 25.076; scale 0.946

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description	Pred. No.
1	56	68.3	473	051898	HYPOTHETICAL 51.9 KD P	1.36e-01
2	52	63.4	389	074894	26S PROTEASE REGULATOR	1.28e+00
3	52	63.4	3410	14	POLYPROTEIN.	1.28e+00
4	52	63.4	3410	14	POLYPROTEIN.	1.28e+00
5	52	63.4	3410	14	POLYPROTEIN PRECURSOR.	1.28e+00
6	52	63.4	3410	14	POLYPROTEIN PRECURSOR.	1.28e+00
7	52	63.4	3411	14	POLYPROTEIN.	1.28e+00
8	52	63.4	3411	14	POLYPROTEIN.	1.28e+00
9	51	62.2	158	2	PREPILIN.	2.20e+00
10	51	62.2	206	11	CYCLIN K (FRAGMENT).	2.20e+00
11	51	62.2	357	4	CYCLIN K.	2.20e+00
12	51	62.2	364	14	MATRIX PROTEIN.	2.20e+00
13	51	62.2	364	14	MATRIX PROTEIN.	2.20e+00
14	51	62.2	364	14	MATRIX PROTEIN.	2.20e+00
15	51	62.2	364	14	MATRIX PROTEIN.	2.20e+00
16	51	62.2	587	4	KIAA0589 PROTEIN (FRAG	2.20e+00
17	50	61.0	202	2	GCHA.	3.76e+00
18	50	61.0	253	5	F52G3.2 PROTEIN.	3.76e+00
19	49	59.8	184	2	GTP CYCLOHYDROLASE I.	5.38e+00
20	49	59.8	205	2	GTP CYCLOHYDROLASE I.	5.38e+00

21	49	59.8	400	2	068995	HYPOTHETICAL 44.2 KD P	5.38e+00
22	49	59.8	584	10	004886	PECTINESTERASE (EC 3.1	5.38e+00
23	49	59.8	584	10	004888	PECTINESTERASE (EC 3.1	5.38e+00
24	49	59.8	513	5	Q23841	PUTATIVE GDH.	5.38e+00
25	49	59.8	764	5	Q24708	CORKSCREW PROTEIN (EC	5.38e+00
26	49	59.8	841	5	Q24032	CORKSCREW PROTEIN Y122	5.38e+00
27	49	59.8	945	5	Q24033	CORKSCREW PROTEIN 4A (5.38e+00
28	49	59.8	1218	4	Q15816	TRANSMEMBRANE PROTEIN	5.38e+00
29	49	59.8	1596	11	062766	SSECKS.	5.38e+00
30	49	59.8	2374	5	045377	F22G12.5 PROTEIN.	5.38e+00
31	48	58.5	144	1	028064	METHYLMALONYL-COA MUTA	1.07e+01
32	48	58.5	194	2	033723	GTP CYCLOHYDROLASE (EC	1.07e+01
33	48	58.5	230	14	Q82681	POLYMERASE-ASSOCIATED	1.07e+01
34	48	58.5	296	2	P74223	GLUTAMINE-BINDING PROT	1.07e+01
35	48	58.5	780	10	004346	REGULATORY PROTEIN VIV	1.07e+01
36	47	57.3	209	14	040596	NEF GENE (STRAIN KUB)	1.07e+01
37	47	57.3	210	3	P87174	PROBABLE DNA-BINDING P	1.79e+01
38	47	57.3	299	2	053342	HYPOTHETICAL 31.6 KD P	1.79e+01
39	47	57.3	341	5	Q20249	F40G12.1 PROTEIN.	1.79e+01
40	47	57.3	364	14	090338	MATRIX PROTEIN.	1.79e+01
41	47	57.3	562	2	P71654	HYPOTHETICAL 58.8 KD P	1.79e+01
42	47	57.3	561	11	070161	PHOSPHATIDYLINOSITOL-4	1.79e+01
43	47	57.3	751	2	033613	CATALASE (EC 1.1.1.1.6)	1.79e+01
44	47	57.3	1105	4	060518	RAN-GTP BINDING PROTEI	1.79e+01
45	47	57.3	2531	5	016004	NOTCH HOMOLOG.	1.79e+01

ALIGNMENTS

RESULT 1
ID 051898 PRELIMINARY: PRT; 473 AA.
AC 051898;
DT 01-JUN-1998 (TREMREL. 06, CREATED)
DT 01-JUN-1998 (TREMREL. 06, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 51.9 KD PROTEIN.
OS BORRELIA AFZELII.
OC BACTERIA; SPIROCHAETALES; SPIROCHAETACEAE; BORRELIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-R-IP3;
RX MEDLINE; 98065594.
RA CASJENS S., MURPHY M., DELANGE M., SAMPSON L., VAN VUGT R.,
RA HUANG W.M.;
RT "Telomeres of the linear chromosomes of Lyme disease spirochaetes:
RT nucleotide sequence and possible exchange with linear plasmid
RT telomeres.";
RL MOL. MICROBIOL. 26:581-596(1997).
DR EMBL; AF008219; G2697115; -
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 473 AA; 51868 MW; BFA2E4DE CRC32;

Query Match 58.3%; Score 56; DB 2; Length 473;
Best Local Similarity 54.5%; Pred. No. 1.36e-01;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
Db 36 QIGDGPKEII 46
QY 2 QIGNDPNRDIL 12
|||: |||:
|||: |||:

RESULT 2
ID 074894 PRELIMINARY: PRT; 389 AA.
AC 074894;
DT 01-NOV-1998 (TREMREL. 08, CREATED)
DT 01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE 26S PROTEASE REGULATORY SUBUNIT 6S HOMOLOG.
GN SPC576.10C.
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
OC SCHIZOSACCHAROMYCETACEAE; SCHIZOSACCHAROMYCETACEAE;
OC SCHIZOSACCHAROMYCES.


```

RP SEQUENCE FROM N.A.
RC STRAIN=VACCINE STRAIN 17DD;
RA GALLER R.;
RL SUBMITTED (NOV-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: U17056; G829367; -.
DR PFAM: PF00869; Flavi_glycoprot; 1.
DR PFAM: PF00948; Flavi_NSI; 1.
DR PFAM: PF00949; Flavi_helicase; 1.
DR PFAM: PF00972; Flavi_NSI; 1.
DR PFAM: PF01002; Flavi_NSI; 1.
DR PFAM: PF01003; Flavi_NSI; 1.
DR PFAM: PF01004; Flavi_M; 1.
DR PFAM: PF01005; Flavi_NSI; 1.
KW POLYPROTEIN.
FT CHAIN 210 284 M PROTEIN.
FT CHAIN 285 777 ENVELOPE PROTEIN.
FT CHAIN 778 1186 NSI.
FT CHAIN 1187 1333 POTENTIAL.
FT CHAIN 1334 1482 POTENTIAL.
FT CHAIN 1483 2105 NS3.
FT CHAIN 2107 2393 POTENTIAL.
FT CHAIN 2394 2505 POTENTIAL.
FT CHAIN 2506 2506 NS5.
SQ SEQUENCE 3410 AA; 379301 MW; 06C3D225 CRC32;

Query Match 53.4%; Score 52; DB 14; Length 3410;
Best Local Similarity 70.0%; Pred. No. 1.28e+00;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 1946 GRIGRNPND 1955
QY 1 GQIGNDPNRD 10

RESULT 6
ID Q89276 PRELIMINARY; PRT; 3410 AA.
AC Q89276;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1998 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE POLYPROTEIN PRECURSOR.
OS YELLOW FEVER VIRUS.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; FLAVIVIRIDAE;
OC FLAVIVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VACCINE STRAIN 17D-213;
RA DOS SANTOS C.N., POST P.R., CARVALHO R., FERREIRA I., RICE C.M.,
RA GALLER R.;
RL VIRUS RES. 35:35-41(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=VACCINE STRAIN 17D-213;
RA GALLER R.;
RL SUBMITTED (NOV-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: U17067; G829369; -.
DR PFAM: PF00869; Flavi_glycoprot; 1.
DR PFAM: PF00948; Flavi_NSI; 1.
DR PFAM: PF00949; Flavi_helicase; 1.
DR PFAM: PF00972; Flavi_NSI; 1.
DR PFAM: PF01002; Flavi_NSI; 1.
DR PFAM: PF01003; Flavi_NSI; 1.
DR PFAM: PF01004; Flavi_M; 1.
DR PFAM: PF01005; Flavi_NSI; 1.
KW POLYPROTEIN.
FT CHAIN 210 284 M PROTEIN.
FT CHAIN 285 777 ENVELOPE PROTEIN.
FT CHAIN 778 1186 NSI.
FT CHAIN 1187 1353 PUTATIVE NON-STRUCTURAL PROTEIN NS2A.
FT CHAIN 1354 1482 PUTATIVE NON-STRUCTURAL PROTEIN NS2B.
FT CHAIN 1483 2105 NS3.
FT CHAIN 2107 2393 PUTATIVE NON-STRUCTURAL PROTEIN NS4A.
FT CHAIN 2394 2505 PUTATIVE NON-STRUCTURAL PROTEIN NS4B.

FT CHAIN 2506 2506 NS5.
SQ SEQUENCE 3410 AA; 379357 MW; D225424D CRC32;

Query Match 63.4%; Score 52; DB 14; Length 3410;
Best Local Similarity 70.0%; Pred. No. 1.28e+00;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 1946 GRIGRNPND 1955
QY 1 GQIGNDPNRD 10

RESULT 7
ID Q91857 PRELIMINARY; PRT; 3411 AA.
AC Q91857;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE POLYPROTEIN.
OS YELLOW FEVER VIRUS.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; FLAVIVIRIDAE;
OC FLAVIVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17D-204-USA VACCINE;
RX MEDLINE; 98376360.
RA XIE H., CASS A.R., BARRETT A.D.T.;
RT "Yellow fever 17D vaccine virus isolated from healthy vaccinees
RT accumulates very few mutations."
RL VIRUS RES. 55:93-99(1998).
DR EMBL: AF052438; G328892; -.
KW POLYPROTEIN.
SQ SEQUENCE 3411 AA; 379584 MW; 1C9F6704 CRC32;

Query Match 63.4%; Score 52; DB 14; Length 3411;
Best Local Similarity 70.0%; Pred. No. 1.28e+00;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 1947 GRIGRNPND 1956
QY 1 GQIGNDPNRD 10

RESULT 8
ID Q98803 PRELIMINARY; PRT; 3411 AA.
AC Q98803;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE POLYPROTEIN.
OS YELLOW FEVER VIRUS.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; FLAVIVIRIDAE;
OC FLAVIVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=85-82H IVORY COAST;
RA PISANO M.R., TOLOU H., NICOLI J.;
RL SUBMITTED (APR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: U54798; G1314775; -.
DR PFAM: PF00869; Flavi_glycoprot; 1.
DR PFAM: PF00948; Flavi_NSI; 1.
DR PFAM: PF00949; Flavi_helicase; 1.
DR PFAM: PF00972; Flavi_NSI; 1.
DR PFAM: PF01002; Flavi_NSI; 1.
DR PFAM: PF01003; Flavi_NSI; 1.
DR PFAM: PF01004; Flavi_M; 1.
DR PFAM: PF01005; Flavi_NSI; 1.
KW POLYPROTEIN.
SQ SEQUENCE 3411 AA; 379170 MW; E3416193 CRC32;

Query Match 63.4%; Score 52; DB 14; Length 3411;
Best Local Similarity 70.0%; Pred. No. 1.28e+00;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

```

Db 1947 GRIGRPNRD 1956
QY 1 GOIGNDPNRD 10

RESULT 9
ID Q95908 PRELIMINARY; PRT; 158 AA.
AC Q95908;
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE PRELIM.
OS MORAXELLA BOVIS.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PSEUDOMONAS GROUP;
OC MORAXELLACEAE; MORAXELLA.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=SW07 / SEROGROUP B;
RX MEDLINE; 94327452.
RA ATWELL J.L., TENNENT J.M., LEPPER A.W., ELLEMAN T.C.;
RT "Characterization of pilin genes from seven serologically defined
RT prototype strains of Moraxella bovis.";
RL J. BACTERIOL. 176:4875-4882(1994).
DR EMBL; L52969; G488318; -.
DR PFAM; PF00114; pilin; 1.
KW METHYLATION.
FT MOD.RES
SQ SEQUENCE 158 AA; 16590 MW; D2841B31 CRC32;

Query Match 62.2%; Score 51; DB 2; Length 158;
Best Local Similarity 66.7%; Pred. No. 2.20e+00;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 109 GRIGRPNRD 120
QY 1 GOIGNDPNRD 12

RESULT 10
ID Q88874 PRELIMINARY; PRT; 206 AA.
AC Q88874;
DT 01-NOV-1998 (TREMREL. 08, CREATED)
DT 01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE CYCLIN K (FRAGMENT).
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=F48E9;
RX MEDLINE; 98298273.
RA EDWARDS M.C., WONG C., ELLEDGE S.J.;
RT "Human cyclin K, a novel RNA polymerase II-associated cyclin
RT possessing both carboxy-terminal domain kinase and Cdk-activating
RT kinase activity.";
RL MOL. CELL. BIOL. 18:4291-4300(1998).
DR EMBL; AF060517; G3746552; -.
FT NON_TER
SQ SEQUENCE 206 AA; 24071 MW; A154D984 CRC32;

Query Match 62.2%; Score 51; DB 11; Length 206;
Best Local Similarity 41.7%; Pred. No. 2.20e+00;
Matches 5; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

Db 114 GQGGDDPKKEVM 125
QY 1 GOIGNDPNRD 12

RESULT 11
ID Q75909 PRELIMINARY; PRT; 357 AA.
AC Q75909;
DT 01-NOV-1998 (TREMREL. 08, CREATED)
DT 01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE CYCLIN K.
GN CPR4.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98298273.
RA EDWARDS M.C., WONG C., ELLEDGE S.J.;
RT "Human cyclin K, a novel RNA polymerase II-associated cyclin
RT possessing both carboxy-terminal domain kinase and Cdk-activating
RT kinase activity.";
RL MOL. CELL. BIOL. 18:4291-4300(1998).
DR EMBL; AF060515; G3746549; -.
SQ SEQUENCE 357 AA; 41293 MW; 547C5CE6 CRC32;

Query Match 62.2%; Score 51; DB 4; Length 357;
Best Local Similarity 41.7%; Pred. No. 2.20e+00;
Matches 5; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

Db 130 GQGGDDPKKEVM 141
QY 1 GOIGNDPNRD 12

RESULT 12
ID Q92330 PRELIMINARY; PRT; 364 AA.
AC Q92330;
DT 01-NOV-1998 (TREMREL. 08, CREATED)
DT 01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE MATRIX PROTEIN.
GN M.
OS NEWCASTLE DISEASE VIRUS STRAIN F48E9.
OC VIRUSES; SSRNA NEGATIVE-STRAND VIRUSES; MONONEGAVIRALES;
OC PARAMYXOVIRIDAE; PARAMYXOVIRINAE; RUBULAVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RA CHEN H., LIU S., CHEN J., ZHANG B., LU J.;
RL SUBMITTED (SEP-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF089819; G3695291; -.
KW MATRIX PROTEIN.
SQ SEQUENCE 364 AA; 39783 MW; B7CE8893 CRC32;

Query Match 62.2%; Score 51; DB 14; Length 364;
Best Local Similarity 41.7%; Pred. No. 2.20e+00;
Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 76 GMINDNPKRELL 87
QY 1 GOIGNDPNRD 12

RESULT 13
ID Q83834 PRELIMINARY; PRT; 364 AA.
AC Q83834;
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE MATRIX PROTEIN.
OS NEWCASTLE DISEASE VIRUS (NDV).
OC VIRUSES; SSRNA NEGATIVE-STRAND VIRUSES; MONONEGAVIRALES;
OC PARAMYXOVIRIDAE; PARAMYXOVIRINAE; RUBULAVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-KIMBER;
RA SEAL B.S.;
RL VIRUS GENES 11:217-224(1996).

```

Job time : 36 secs.

DR EMBL; U25831; G1016659; -.
DR PFAM; PF00661; Matrix; 1.
KW MATRIX PROTEIN.
SQ SEQUENCE 364 AA; 39723 MW; F0E2A4AB CRC32;

Query Match 62.2%; Score 51; DB 14; Length 364;
Best Local Similarity 41.7%; Pred. No. 2.20e+00;
Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 76 GMINDNPKRELL 87
| | : : : : :
QY 1 GOIGNDPNRDIL 12

RESULT 14
ID Q83838 PRELIMINARY; PRT; 364 AA.
AC Q83838;
DT 01-NOV-1996 (TREMELREL. 01, CREATED)
DT 01-NOV-1996 (TREMELREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMELREL. 08, LAST ANNOTATION UPDATE)
DE MATRIX PROTEIN.
OS NEWCASTLE DISEASE VIRUS (NDV).
OC VIRUSES; SSRNA NEGATIVE-STRAND VIRUSES; MONONEGAVIRALES;
OC PARAMYXOVIRIDAE; PARAMYXOVIRINAE; RUBULAVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TEXAS/GB;
RA SEAL B.S.;
RL VIRUS GENES 11:217-224(1996).
DR EMBL; U25835; G1016667; -.
DR PFAM; PF00561; Matrix; 1.
KW MATRIX PROTEIN.
SQ SEQUENCE 364 AA; 39464 MW; 72A78A51 CRC32;

Query Match 62.2%; Score 51; DB 14; Length 364;
Best Local Similarity 41.7%; Pred. No. 2.20e+00;
Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 76 GMINDNPKRELL 87
| | : : : : :
QY 1 GOIGNDPNRDIL 12

RESULT 15
ID Q83840 PRELIMINARY; PRT; 364 AA.
AC Q83840;
DT 01-NOV-1996 (TREMELREL. 01, CREATED)
DT 01-NOV-1996 (TREMELREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMELREL. 08, LAST ANNOTATION UPDATE)
DE MATRIX PROTEIN.
OS NEWCASTLE DISEASE VIRUS (NDV).
OC VIRUSES; SSRNA NEGATIVE-STRAND VIRUSES; MONONEGAVIRALES;
OC PARAMYXOVIRIDAE; PARAMYXOVIRINAE; RUBULAVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ULSTER;
RA SEAL B.S.;
RL VIRUS GENES 11:217-224(1996).
DR EMBL; U25837; G1016671; -.
DR PFAM; PF00661; Matrix; 1.
KW MATRIX PROTEIN.
SQ SEQUENCE 364 AA; 39645 MW; 4F9287D2 CRC32;

Query Match 62.2%; Score 51; DB 14; Length 364;
Best Local Similarity 41.7%; Pred. No. 2.20e+00;
Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 76 GMINDNPKRELL 87
| | : : : : :
QY 1 GOIGNDPNRDIL 12

This page blank (uspto)

W P S R L H
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Tue Aug 17 15:57:49 1999; MasPar time 3.90 Seconds
Tabular output not generated. 65.384 Million cell updates/sec

Title: >US-09-049-847-3
Description: (1-12) from US09049847.pap
Perfect Score: 82
Sequence: 1 GOIGNDPNRDIL 12
Scoring table: PAM 150
Gap 15

Searched: 170751 seqs, 21266608 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 17.007; Variance 45.283; scale 0.367

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					
Result No.	Score	Query Match	Length DB	ID	Pred. No.
1	82	100.0	12 36	W67035	Tetanus toxin fragmen
2	82	100.0	12 35	W78835	Tetanus toxoid protei
3	82	100.0	13 30	W50109	Pan DR binding peptid
4	82	100.0	452 2	R12471	Tetanus toxin fragmen
5	82	100.0	573 2	P70345	Portion of B fragment
6	82	100.0	618 33	W48909	SOD-1/TRC hybrid prot
7	78	95.1	12 1	R06308	Tetanus toxin epitope
8	51	62.2	401 25	W33109	Streptomyces sp. lam
9	51	62.2	548 4	R22197	Sequence encoded by b
10	50	61.0	522 35	W69726	Saccharomyces cerevis
11	50	61.0	526 5	R36174	Protein disulphide is
12	49	59.8	220 24	W12662	Orange pectin methyl
13	49	59.8	584 37	W29964	Aspergillus aculeatus
14	49	59.8	584 24	W12661	Orange pectin methyl
15	49	59.8	1218 19	W05833	Human serrate-1 (HJ1)
16	49	59.8	1346 30	W31346	Rat tumour suppressor

17	49	59.8	1596 30	W31347	Rat tumour suppressor
18	48	58.5	57 38	W83767	Polypeptide fragment
19	47	57.3	821 25	W26664	Yeast checkpoint cont
20	46	56.1	208 34	W38500	S. pneumoniae 50S rib
21	46	56.1	306 15	R75132	Porcine carboxypeptid
22	46	56.1	329 20	W06174	Mature HCPB with PelB
23	46	56.1	349 21	W13763	Carboxypeptidase B mu
24	46	56.1	349 21	W13761	Carboxypeptidase B mu
25	46	56.1	349 20	W06182	Modified HCPB (D253R)
26	46	56.1	349 21	W13759	Carboxypeptidase B mu
27	46	56.1	349 21	W13762	Carboxypeptidase B mu
28	46	56.1	349 21	W13750	Carboxypeptidase B mu
29	46	56.1	349 21	W13765	Carboxypeptidase B mu
30	46	56.1	349 21	W13752	Carboxypeptidase B mu
31	46	56.1	349 21	W13757	Carboxypeptidase B mu
32	46	56.1	349 20	W06173	Carboxypeptidase B mu
33	46	56.1	349 21	W13764	Carboxypeptidase B mu
34	46	56.1	349 21	W13751	Carboxypeptidase B mu
35	46	56.1	349 21	W13758	Carboxypeptidase B mu
36	46	56.1	349 21	W13753	Carboxypeptidase B mu
37	46	56.1	349 21	W13760	Carboxypeptidase B mu
38	46	56.1	349 20	W06181	Modified HCPB (D253K)
39	46	56.1	404 15	R75131	Porcine Tvr-His-Met p
40	46	56.1	415 20	W06172	Human pancreatic carb
41	46	56.1	415 35	W74476	Human pancreatic carb
42	46	56.1	424 20	W06175	ProHCPB with PelB lea
43	46	56.1	613 35	W74481	Human pancreatic carb
44	46	56.1	655 4	R20139	Sequence encoded by n
45	46	56.1	715 4	R20138	Sequence encoded by p

ALIGNMENTS

RESULT 1
ID W67035 standard; peptide; 12 AA.

AC W67035;
DT 15-DEC-1998 (first entry)
DE Tetanus toxin fragment (residues 1273-1284).
KW Tetanus toxin; vaccine; antibody; carbohydrate peptide conjugate;
KW dendritic poly-lysine; epitope; tumour.
OS Clostridium tetani.
PN W09843677-A1.
PD 08-OCT-1998.
PF 27-MAR-1998; E01922.
PR 27-MAR-1997; US-041726.
PA (INSP) INST PASTEUR.
PI Bay S, Cantacuzene D, Leclerc C, Lo-man R;
DR WPI, 98-557071/47.
PT Carbohydrate peptide conjugate used as vaccine - comprises carrier
PT with dendritic poly-lysine enabling multiple epitopes to be
PT covalently attached
PS Disclosure; Page 13; 55pp; English.
CC The invention relates to a new carbohydrate peptide conjugate, which
CC comprises a carrier with a dendritic poly-lysine enabling multiple
CC epitopes to be covalently attached to it. Also claimed are: (1) an
CC antibody purified from biological fluid or cells of organisms
CC administered with the carbohydrate peptide conjugate, and (2) a diagnosis
CC kit comprising antigen-specific antibodies elicited by immunisation with
CC the carbohydrate peptide conjugate. The peptide conjugate, antibody and
CC diagnosis kit are used to provide pharmaceutical compositions and
CC vaccines against tumours. These can be used to support an immune response
CC against viral infections caused by hepatitis virus, HIV or cytomegalo
CC virus. They can be used to enhance immune responses, especially B- and T-
CC cell responses, of humans and animals against bacterial infections. The
CC carbohydrate peptide conjugate stimulates the antibody and T-cell
CC response without stimulating undesired immune responses. The composition
CC is capable of increasing the survival of tumour bearing humans and
CC animals. The present sequence corresponds to residues 1273-1284 of
CC tetanus toxin. The synthetic peptide corresponding to this sequence may
CC be used as an epitope in a carbohydrate peptide conjugate.
SQ Sequence 12 AA;

Query Match 100.0%; Score 82; DB 36; Length 12;

Best Local Similarity 100.0%; Pred. No. 5.29e-03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 gqigndpnrtil 12
QY 1 GQIGNDPNRDIL 12

RESULT 2
ID W78835 standard; peptide; 12 AA.
AC W78835;
DE Tetanus toxin protein fragment 1273-1284.
KW Microparticle; delivery; polymeric matrix; autoantigen; tumour antigen;
KW Class II associated peptide; pathogen; gene therapy; genetic disease;
KW infection; downregulation; immune response.
OS Synthetic.
PN W09831398-AL.
PD 23-JUL-1998.
PF 22-JAN-1998; U01499.
PR 06-JAN-1998; US-003253.
PR 22-JAN-1997; US-787547.
PA (PANG-) PANCAEA PHARM INC.
PI Curley JM, Hedley ML, Langer RS, Lunsford LB;
FI WPI; 98-427556/36.
PT New preparations of microparticles - comprising a synthetic polymer
PT matrix and nucleic acid comprising an expression vector for use in
PT gene therapy
PS Disclosure; Page 8; 101pp; English.
CC A microparticle preparation (MP) has been developed, consisting of
CC microparticles having a diameter of less than 100 nm. The MP comprises:
CC (a) a polymeric matrix (PM) consisting of one or more synthetic polymers
CC having a solubility in water of less than 1 mg/l; and (b) an expression
CC vector selected from RNA molecules (at least 50% of which are closed
CC circles) or circular plasmid DNA (at least 50% of which are supercoiled).
CC Also described is a MP of at most 20 microns in diameter, comprising: (a)
CC a PM; and (b) a NAM comprising an expression control sequence operatively
CC linked to a coding sequence, where the coding sequence encodes an
CC expression product selected from: (1) a polypeptide at least 7 amino
CC acids in length, having a sequence identical to the sequence of: (i) a
CC fragment of a naturally-occurring mammalian protein; or (ii) a fragment
CC of a naturally-occurring protein from an infectious agent which infects
CC a mammal; (2) a peptide having a length and sequence which permits it to
CC bind to an MHC class I or II molecule; and (3) the polypeptide or the
CC peptide linked to a trafficking sequence. W69763 to W69765, and W78793
CC to W78897 are peptide fragments for use in the present invention. The
CC MMs are highly effective vehicles for the delivery of polynucleotides
CC into phagocytic cells. They can be used for gene therapy, e.g. for
CC treating genetic diseases, infections or tumours or for downregulating
CC an immune response.
SQ Sequence 12 AA;

Query Match 100.0%; Score 82; DB 35; Length 12;
Best Local Similarity 100.0%; Pred. No. 5.29e-03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 gqigndpnrtil 12
QY 1 GQIGNDPNRDIL 12

RESULT 3
ID W50109 standard; peptide; 13 AA.
AC W50109;
DE 30-JUN-1998 (first entry)
DE Pan DR binding peptide (6).
DE Pan DR binding peptide; antigen binding site; MHC molecule;
KW DR locus.
OS Synthetic.
PN US5736142-A.
PD 07-APR-1998.
PF 14-SEP-1994; 305871.
PR 14-SEP-1994; US-305871.

PR 14-SEP-1993; US-121101.
PA (CYTE-) CYTEL CORP.
PI Alexander JL, Gaeta F, Grey HM, Sette A, Sidney J;
DR WPI; 98-239154/21.
PT Peptides that bind to MHC molecules of all DR alleles - inhibiting
PT or inducing MHC Class II mediated activation of T cells
PS Disclosure; Columns 33-34; 29pp; English.
CC The present sequence, a pan DR binding peptide, is capable of
CC binding antigen binding sites on MHC molecules, which are encoded
CC by most of the alleles of a DR locus. The peptide can be used to
CC inhibit or induce MHC Class II mediated activation of T-cells or
CC helper T-cells, which themselves mediate a CTL response. The
CC peptide can be used in mammals, especially humans, to inhibit
CC T-cell-mediated events involved in allograft rejection, allergic
CC responses and autoimmunity and as a vaccine adjuvant for enhancing
CC an immune response against an administered immunogen. The peptide
CC can be used with other immunogens to treat, e.g. prostate cancer,
CC hepatitis B, hepatitis C, AIDS, renal and cervical carcinoma,
CC lymphoma, CMV and condyloma acuminatum.
SQ Sequence 13 AA;

Query Match 100.0%; Score 82; DB 30; Length 13;
Best Local Similarity 100.0%; Pred. No. 5.29e-03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2 gqigndpnrtil 13
QY 1 GQIGNDPNRDIL 12

RESULT 4
ID R12471 standard; Protein; 452 AA.
AC R12471;
DT 05-AUG-1991 (first entry)
DE Tetanus toxin fragment C encoded by gene with increased G+C content.
DE Terminator; vaccine.
KW Terminator; vaccine.
OS Synthetic.
PN EP-430645-A.
PD 05-JUN-1991.
PF 27-NOV-1990; 312870.
PR 28-NOV-1989; GB-026832.
PR 17-MAR-1990; GB-006097.
PA (WELL) WELLCOME FOUNDATION LTD.
PI Makoff AJ, Romanos MA, Clare JJ, Fairweather NF;
DR WPI; 91-166115/23.
DR N-PSDB; Q12121.
PT DNA sequence encoding tetanus toxin fragment C - useful in the
PT manufacture of vaccines for immunity to tetanus utilising yeast
PT as host organism.
PS Disclosure; Fig 2; 50pp; English.
CC The (G+C) content of the synthetic gene is increased by 47% wrt the
CC native sequence. This eliminates six "terminator" regions which
CC were found to be present in (A+T) rich regions. The terminators
CC (termination/endo-nucleolytic processing/polyadenylation sites)
CC were previously responsible for incomplete transcription of the
CC mRNA. The elimination of these elements (using codon degeneracy)
CC provided for successful expression in yeast of the tetanus toxin
CC fragment C.
SQ Sequence 452 AA;

Query Match 100.0%; Score 82; DB 2; Length 452;
Best Local Similarity 100.0%; Pred. No. 5.29e-03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 410 gqigndpnrtil 421
QY 1 GQIGNDPNRDIL 12

RESULT 5
ID P70345 standard; Protein; 573 AA.
AC P70345;
DT 22-APR-1991 (first entry)

DE Portion of B fragment and all of the C fragment of tetanus toxin.
 KW TT; vaccine.
 OS Clostridium tetani.
 PN EP-209281-A.
 PD 21-JAN-1987.
 PF 27-JUN-1986; 305029.
 PR 28-JUN-1985; GB-016442.
 PA (WELL) Wellcome Foundation Ltd.
 PI Fairweather NF;
 DR WPI; 87-015995/03.
 DR N-PSDB; N70545.
 PI Cloned DNA sequence coding for tetanus toxin - or its fragments
 PT contg. epitope used to express antigens for vaccine production.
 PS Claim 4; Fig 1; 3pp; English.
 CC Gene product comprises a tetanus toxin fragment, which may be
 CC expressed in a transformed host, and used as an antigen in vaccine
 CC production, against the disease.
 SQ Sequence 573 AA;

Query Match 100.0%; Score 82; DB 2; Length 573;
 Best Local Similarity 100.0%; Pred. No. 5.29e-03;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 531 gqigndpnrdil 542
 QY 1 GQIGNDPNRDIL 12
 |||||

RESULT 6
 ID W48909 standard; Protein; 618 AA.
 AC W48909.
 DT 23-SEP-1998 (first entry)
 DE SOD-1/TTC hybrid protein.
 KW Chimeric; copper-zinc superoxide dismutase; SOD-1; TTC; SOD:Tet451;
 KW tetanus toxin fragment C; tetanus holotoxin; nerve cell; stroke;
 KW neurological disorder; oxidative stress; brain hypoxia-reperfusion;
 KW epilepsy; Parkinson's disease; Huntington's disease.
 OS Chimeric - Homo sapiens.
 OS Chimeric - Clostridium tetani.
 FH Key Location/Qualifiers
 FT Region 1..163
 FT /note= "SOD-1"
 FT Region 168..618
 FT /note= "TTC moiety"

PN US780024-A.
 PD 14-JUL-1998.
 PF 21-JUN-1996; 668381.
 PR 23-JUN-1995; US-000473.
 PR 21-JUN-1996; US-668381.
 PA (GENO) GEN HOSPITAL CORP.
 PA (UYMA-) UNIV MARYLAND BALTIMORE.
 PI Brown RH, Fishman PS, Francis JW, Hosler BA;
 DR WPI; 98-412995/35.
 DR N-PSDB; V32580.
 PT New hybrid protein of superoxide dismutase and tetanus toxin
 PT fragment C - having increased uptake by neurons and retention of
 PT enzymatic activity in these cells, for treating neurological
 PT diseases associated with oxidative stress
 PS Claim 7; Columns 23-26; 23pp; English.
 CC The present sequence represents an enzymatically active human
 CC copper-zinc superoxide dismutase (SOD-1) fused at its carboxyl
 CC terminus with the tetanus toxin fragment C (TTC) moiety. The TTC
 CC moiety constitutes amino acid residues 865-1315 of the tetanus
 CC holotoxin. The hybrid protein, referred as SOD:Tet451, is claimed
 CC to have the following properties: (a) it exhibits Cu/Zn SOD enzymatic
 CC activity; (b) the TTC moiety selectively binds to nerve cells and
 CC allows uptake of the hybrid protein into these cells; and (c) it
 CC retains substantial SOD enzymatic activity following cellular uptake.
 CC SOD:Tet451 is claimed to be useful for treating neurological disorders
 CC associated with oxidative stress, e.g. stroke, brain hypoxia-reperfusion,
 CC epilepsy, Parkinson's and Huntington's diseases.
 SQ Sequence 618 AA;

Query Match 100.0%; Score 82; DB 33; Length 618;
 Best Local Similarity 100.0%; Pred. No. 5.29e-03;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 576 gqigndpnrdil 587
 QY 1 GQIGNDPNRDIL 12
 |||||

RESULT 7
 ID R06308 standard; peptide; 12 AA.
 AC R06308;
 DT 04-DEC-1990 (first entry)
 DE Tetanus toxin epitope.
 KW Tetanus toxin; vaccine; major histocompatibility complex; MHC;
 KW antimalarial.
 OS Synthetic.
 PN EP-378881-A.
 PD 25-JUL-1990.
 PF 27-DEC-1989; 203318.
 PR 17-JAN-1989; IT-019110.
 PR 16-NOV-1989; IT-022409.
 PA (ENIE) ENRIGERCHE SPA.
 PI Pessi A, Bianchi E, Verdini AS, Corradin G;
 DR WPI; 90-225582/30.
 PT Synthetic peptide(s) corresp. to tetanus toxin epitope(s) - used
 PT as universal carriers for prepn. of immunogenic conjugate(s) for
 PT use as vaccines.
 PS Claim 1; Page 17; 20pp; English.
 CC Epitopic peptides may be used with synthetic hapten derived from
 CC a pathogen to generate an immune response to the pathogen.
 CC Peptides are recognised by numerous T-helper cell clones within
 CC the context of a wide range of alleles of the human MHC.
 CC The peptides may be used in an antimalarial vaccine inducing Ab.
 CC response to P.falciparum.
 SQ Sequence 12 AA;

Query Match 95.1%; Score 78; DB 1; Length 12;
 Best Local Similarity 91.7%; Pred. No. 1.79e-02;
 Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 1 geigndpnrdil 12
 QY 1 GQIGNDPNRDIL 12
 |:|||||

RESULT 8
 ID W33109 standard; Protein; 401 AA.
 AC W33109;
 DT 28-JAN-1998 (first entry)
 DE Streptomyces sp. laminary pentose forming enzyme.
 KW Laminary pentose forming enzyme; recombinant production.
 OS Streptomyces sp.
 FH Key Location/Qualifiers
 FT Peptide 1..35
 FT /label= sig_peptide
 FT Peptide 36..401
 FT /label= mat_peptide
 PN J09262090-A.
 PD 07-OCT-1997.
 PF 28-MAR-1996; 074227.
 PR 28-MAR-1996; JP-074227.
 PA (DNTN) DAINIPPON INK & CHEM INC.
 DR WPI; 97-544153/50.
 DR N-PSDB; T88370.
 PT DNA encoding laminary pentose forming enzyme - useful for large
 PT scale recombinant production
 PS Claim 1; Pages 7-9; 10pp; Japanese.
 CC The DNA encoding the present sequence, Streptomyces sp. DIC-108
 CC laminary pentose forming enzyme, is useful for the large scale
 CC recombinant production of the enzyme.
 SQ Sequence 401 AA;

Query Match 62.2%; Score 51; DB 25; Length 401;
 Best Local Similarity 72.7%; Pred. No. 4.37e+01;
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 139 qnptdprdil 149
 | :|||||
 QY 2 QIGNDPNRDIL 12

RESULT 9

ID R22197 standard; Protein; 548 AA.

AC R22197; 24-JUL-1992 (first entry)

DE Sequence encoded by beta-1, 3-glucanase gene.

KW Recombinant enzyme; glucanase; pUV5-G1S.

OS Oerskovia xanthineolytica.

PH Key Location/Qualifiers

FT Peptide 1..36

FT /label= signal

FT region 444..448

FT /label= repeat

FT region 456..474

FT /label= repeat

FT region 488..492

FT /label= repeat

FT region 501..519

FT /label= repeat

PN W09203557-A.

PD 05-MAR-1992.

PR 16-AUG-1991; CA0294.

PR 17-AUG-1990; US-568869.

PA (CANADA) NAT RES COUNCIL CAN.

PI Shen SH, Chretien P, Bastien L, Slliaty SN;

DR WPI; 92-096900/12.

DR N-PSDB: Q22957.

PT Recombinant beta-1,3-glucanase enzyme - produced by transforming

PT E. coli cells with expression vectors and free of protease

PT activity

PS Example; Fig 2; 32pp; English.

CC The inventors claim a recombinant enzyme prepn. comprising beta-1,

CC 3-glucanase or a mutant or variant and a vector contg. DNA encoding

CC it under the control of an exogenous promoter pref. lac UV5. The

CC example uses the native beta-1, 3-glucanase gene isolated from

CC Oerskovia xanthineolytica genomic DNA and expressed in pUV5-G1S.

SQ Sequence 548 AA;

Query Match 62.2%; Score 51; DB 4; Length 548;

Best Local Similarity 63.6%; Pred. No. 4.37e+01;

Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 138 qnadsprnil 148
 | :|||||
 QY 2 QIGNDPNRDIL 12

RESULT 10

ID W69726 standard; Protein; 522 AA.

AC W69726;

DT 16-NOV-1998 (first entry)

DE Saccharomyces cerevisiae protein disulphide isomerase.

KW Saccharomyces cerevisiae; yeast; protein disulphide isomerase; PDI;

KW recombinant; endoplasmic reticulum; localisation signal;

KW food processing.

OS Saccharomyces cerevisiae.

PN W09835049-A1.

PD 13-AUG-1998.

PR 06-FEB-1998; J00498.

PR 07-FEB-1997; JP-038588.

PA (ORIX) ORIENTAL YEAST CO LTD.

PI Ishii N, Matuo Y, Suzuki Y, Tanaka H, Uchida K;

DR WPI; 98-447246/38.

DR N-PSDB: V50470.

PT Active recombinant yeast protein disulphide isomerase - lacking

PT localisation signal, can be produced in quantity for use in food
 PT processing
 PS Disclosure; Fig 2-4; 48pp; Japanese.
 CC The present sequence represents yeast (Saccharomyces cerevisiae)
 CC protein disulphide isomerase (PDI). The present invention describes a
 CC biologically active recombinant yeast PDI, in which the endoplasmic
 CC reticulum localisation signal (His-Asp-Glu-Leu) at the C-terminal is
 CC defective or lacking, is produced by: (a) modifying a gene encoding the
 CC natural PDI by adding, deleting or substituting one or more bases in the
 CC region encoding the endoplasmic reticulum localisation signal so that
 CC part of all of the localisation signal falls; (b) inserting this
 CC modified gene into a suitable vector (such as the yeast expression
 CC vector YEpIG11); (c) transforming a host cell with the vector; (d)
 CC culturing the transformant at a pH near to neutrality (preferably at
 CC pH 6.5 to 8.0), and (e) isolating the PDI from the culture medium. PDI
 CC is used in the processing of foods such as ham, sausages, fish products
 CC and bean-curd (tofu). The modified PDI retains its biological activity
 CC but is expressed into the culture medium from where it can be isolated
 CC by simple purification methods in high yield.
 SQ Sequence 522 AA;

Query Match 61.0%; Score 50; DB 35; Length 522;

Best Local Similarity 54.5%; Pred. No. 5.72e+01;

Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 388 eivndpkkdvl 398
 | :|||||
 QY 2 QIGNDPNRDIL 12

RESULT 11

ID R26174 standard; Protein; 526 AA.

AC R26174;

DT 04-FEB-1993 (first entry)

DE Protein disulphide isomerase.

KW yeast; isomerisation; catalyst; enzyme; disulphide bonds.

OS Saccharomyces cerevisiae.

PN J04197176-A.

PD 16-JUL-1992.

PR 28-NOV-1990; 322620.

PR 28-NOV-1990; JP-322620.

PA (TOFU) TONEN CORP.

DR WPI; 92-288434/35.

DR N-PSDB: Q27388.

PT Yeast protein di:sulphide isomerase - used to give active

PT conformation to protein having inadequate di:sulphide bond

PS Disclosure; Fig 15; 22pp; Japanese.

CC The sequence given is a yeast protein disulphide isomerase (PDI)

CC which catalyses the isomerisation of disulphide bonds. It has a

CC molecular weight of approx 70 kD and has an optimum pH for activity

CC of 8.75. This enzyme can be used to give active conformation to a

CC protein having inadequate disulphide bonds. The gene encoding this

CC protein can be used to produce expression vectors for the large scale

CC production of PDI.

SQ Sequence 526 AA;

Query Match 61.0%; Score 50; DB 5; Length 526;

Best Local Similarity 54.5%; Pred. No. 5.72e+01;

Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 388 eivndpkkdvl 398
 | :|||||
 QY 2 QIGNDPNRDIL 12

RESULT 12

ID W12662 standard; Protein; 220 AA.

AC W12662;

DT 13-NOV-1997 (first entry)

DE Orange pectin methyltransferase isoform L extension sequence.

KW Pectin methyltransferase; PME; orange; de-esterified pectin;

KW esterification; fruit juice; yoghurt; whey protein; casein;

KW protein stabilisation.

Citrus sinensis.
 PN WO9703574-A1.
 PD 06-FEB-1997.
 PF 12-JUL-1996; E03051.
 PR 14-JUL-1995; GB-014438.
 PA (DANI-) DANISCO AS.
 PI Buchholt HC, Christensen TMIE, Kreiberg JD, Nielsen J;
 PI Rasmussen P, Thorsøe H;
 DR NPI; 97-145190/13.
 DR N-PSDB; T51740.
 PT Stabilising protein in acidic soln. - by adding enzymatically
 PT de-esterified pectin, produced using pectin methylesterase
 PS Claim 44; Page 75; 114pp; English.
 CC This 220-amino acid polypeptide comprises an N-terminal extension
 CC sequence of orange pectin methyltransferase (PME). The extension
 CC is located N-terminally to the signal peptide of the PME (W12661)
 CC encoded by PME cDNA clone p034 (T51739). The 64 kDa longer (L)
 CC isoform of PME is believed to be more heat stable than the 36 kDa
 CC shorter (S) isoform. Isoform S is believed to start the initial
 CC de-esterification of a pectin and is then superseded by isoform L.
 CC The 220-amino acid extension can be used to impart or increasing
 CC heat stability to a protein.
 SQ Sequence 220 AA;
 SQ
 Query Match 59.8%; Score 49; DB 24; Length 220;
 Best Local Similarity 50.0%; Pred. No. 7.49e+01;
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
 Db 12 gdnqnepphail 23
 QY 1 GQIGNDPNRDIL 12
 RESULT 13
 ID W72964 standard; Protein; 584 AA.
 AC W72964;
 DT 21-JAN-1999 (first entry)
 DE Aspergillus aculeatus pectin methyl esterase.
 KW Orange; pectin methyl esterase; PME; Aspergillus; foodstuff; marmalade;
 KW fruit; vegetable; jam; preserve.
 OS Aspergillus aculeatus.
 PN WO9847391-A1.
 PD 29-OCT-1998.
 PF 24-APR-1997; IB0673.
 PR 24-APR-1997; GB-008278.
 PA (DANI-) DANISCO AS.
 PI Christensen TMIE, Hyttel S, Kreiberg JD;
 DR NPI; 98-583317/49.
 DR N-PSDB; V64074.
 PT New composition containing pectin methyl esterase and two substrates
 PT - induces gelling in foodstuffs having less than 50% soluble solids
 PT content, used for, e.g. preparation of foodstuff from fruits
 PS Disclosure; Page 44; 29pp; English.
 CC A composition has been developed which comprises a pectin methyl
 CC esterase (PME), and two PME substrates either one of which does not
 CC originate in situ from the other. The present sequence represents a PME
 CC from Aspergillus aculeatus. Also described in the present invention are:
 CC (1) a method of preparing the above composition by forming a mixture of
 CC a PME and two PME substrates either one of which does not originate in
 CC situ from the other; (2) a process comprising adding to a PME substrate,
 CC a PME and another PME substrate, where neither PME substrate originates
 CC in situ from the other; (3) a process of imparting stability to a
 CC reaction medium comprising a PME substrate, by adding at least 1 PME and
 CC a second PME substrate, where neither PME substrate originates in situ
 CC from the other; (4) an aqueous system being in a solidified gel state
 CC and having a soluble solids content of less than 50% w/w, where the
 CC gelling has occurred by use of a high ester PME substrate; and (5) a
 CC foodstuff comprising or prepared from or by the invention. The methods
 CC and products are used in the preparation of foodstuffs from fruit or
 CC vegetable materials containing pectin, such as jams and preserves. The
 CC process is advantageous over prior art in that it comprises addition of
 CC a second PME substrate which will overcome problems associated with
 CC differing quantities and qualities of PME substrates.

SQ Sequence 584 AA;
 Query Match 59.8%; Score 49; DB 37; Length 584;
 Best Local Similarity 50.0%; Pred. No. 7.49e+01;
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
 Db 58 gdnqnepphail 69
 QY 1 GQIGNDPNRDIL 12
 RESULT 14
 ID W12661 standard; Protein; 584 AA.
 AC W12661;
 DT 13-NOV-1997 (first entry)
 DE Orange pectin methylesterase (S isoform).
 KW Pectin methylesterase; PME; orange; de-esterified pectin;
 KW esterification; fruit juice; yoghurt; whey protein; casein;
 KW protein stabilisation.
 OS Citrus sinensis.
 FH Key: Location/Qualifiers
 FT peptide 1..46
 FT /label= Sig_peptide
 FT 47..584
 FT /label= Mat_protein
 PN WO9703574-A1.
 PD 06-FEB-1997.
 PF 12-JUL-1996; E03051.
 PR 14-JUL-1995; GB-014438.
 PA (DANI-) DANISCO AS.
 PI Buchholt HC, Christensen TMIE, Kreiberg JD, Nielsen J;
 PI Rasmussen P, Thorsøe H;
 DR NPI; 97-145190/13.
 DR N-PSDB; T51739.
 PT Stabilising protein in acidic soln. - by adding enzymatically
 PT de-esterified pectin, produced using pectin methylesterase
 PS Claim 18; Page 73; 114pp; English.
 CC A polypeptide comprises a novel orange pectin methylesterase (PME)
 CC that can be obtained by expression of the PME coding sequence
 CC contained in NCIMB 40750 or of a cDNA clone (T51739) obtained from
 CC an orange fruit cDNA library. A longer isoform (see W12662) of the
 CC PME was also identified. Recombinant PME can be expressed in
 CC transformed host cells or transgenic organisms, and used in a
 CC claimed method for de-esterifying a pectin. The de-esterified
 CC pectin is used in claimed methods for stabilising a protein in an
 CC acidic environment (such as fruit juice, drinking yoghurt or drinks
 CC containing whey or milk proteins) without adversely affecting the
 CC viscosity of that environment. The recombinant enzyme can also be
 CC used to esterify pectins in non-aqueous medium. Another novel PME
 CC (see W12660) is also claimed.
 SQ Sequence 584 AA;
 SQ
 Query Match 59.8%; Score 49; DB 24; Length 584;
 Best Local Similarity 50.0%; Pred. No. 7.49e+01;
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
 Db 58 gdnqnepphail 69
 QY 1 GQIGNDPNRDIL 12
 RESULT 15
 ID W05833 standard; Protein; 1218 AA.
 AC W05833;
 DT 28-JAN-1997 (first entry)
 DE Human Serrate-1 (HJ1).
 KW Serrate-1; human jagged-1; HJ1; Notch; cell differentiation;
 KW cell fate; central nervous system; cancer; tissue repair; therapy;
 KW diagnosis; antibody.
 OS Homo sapiens.
 FH Key: Location/Qualifiers
 FT domain 1..1067
 FT /label= Extracellular_domain

FT peptide 14...29
FT /label= Sig_peptide
FT domain 185...229
FT /label= DSL
FT /note= "region of homology with Drosophila Delta
FT and Serrate, predicted to mediate binding
FT with Notch"
FT 234...896
FT domain
FT /label= ELR
FT /note= "epidermal growth factor-like repeat domain"
FT 234...264
FT /label= ELR1
FT 265...299
FT /label= ELR2
FT 300...339
FT /label= ELR3
FT 340...377
FT /label= ELR4
FT 378...415
FT /label= ELR5
FT 416...453
FT /label= ELR6
FT 454...490
FT /label= ELR7
FT 491...528
FT /label= ELR8
FT 529...566
FT /label= ELR9
FT 567...598
FT /label= Partial_ELR
FT 599...632
FT /label= Partial_ELR
FT 633...670
FT /label= ELR10
FT 671...708
FT /label= ELR11
FT 709...747
FT /label= ELR12
FT 748...785
FT /label= ELR13
FT 786...823
FT /label= ELR14
FT 824...862
FT /label= ELR15
FT 863...879
FT /label= Partial_ELR
FT 880...896
FT /label= Partial_ELR
FT 1068...1089
FT domain
FT /label= Transmembrane_domain
FT 1090...1218
FT /label= Intracellular_domain
PN W09627610-A1.
PD 12-SEP-1996
PF 07-MAR-1996; U03172.
PR 07-MAR-1995; US-400159.
PA (IMCR) IMPERIAL CANCER RES TECHNOLOGY.
PA (UYVA) UNIV YALE.
PI Artavanis-Tsakonas S, Gray GE, Henrique DMP, Ish-Horowicz D;
PI Lewis JH, Mann RS, Myat AM;
DR WPI: 96-425379/42.
DR N-PSDB; T40090.
PT Vertebrate Serrate protein and related DNA - used to treat or
PT prevent malignancies characterised by increased Notch activity.
PS Claim 4; Page 95-98; 161pp; English.
CC Human Serrate-1 (W05833) and human Serrate-2 (W05833) are ligands
CC for the zygotic neurogenic locus Notch, and are believed to play a
CC major role in determining cell fates (differentiation) in the
CC central nervous system. Their amino acid sequences were deduced
CC from cDNA clones (see also T40090-91) isolated from human foetal
CC brain cDNA libraries. The proteins, antibodies raised to them,
CC and encoding nucleic acids can be used in the detection of
CC Serrate sequences and in the treatment of disorders of cell fate

CC or differentiation, partic. cancer, nervous system disorders
CC and in tissue repair or regeneration.
SQ Sequence 1218 AA;

Query Match 59.8%; Score 49; DB 19; Length 1218;
Best local Similarity 77.8%; Pred. NO. 7.49e+01;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 114 gndparivl 122
QY 4 GNDPARDIL 12
|||||:|

Search completed: Tue Aug 17 15:58:12 1999
Job time : 23 secs.

M P E R L H
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein ~ protein database search, using Smith-Waterman algorithm
Run on: Tue Aug 17 16:00:27 1999; MasPar time 1.53 Seconds
Tabular output not generated. 79.635 Million cell updates/sec

Title: >US-09-049-847-3
Description: (1-12) from US09049847.pep
Perfect Score: 82
Sequence: 1 GQIGNPNRDIL 12

Scoring table: PAM 150
Gap 15

Searched: 106580 seqs, 10152877 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued

Statistics: Mean 16.008; Variance 42.182; scale 0.379

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	82	100.0	12	2	US-08-787- Sequence 46, Applicati	1.05e-03
2	82	100.0	13	1	US-08-305- Sequence 6, Applicatio	1.05e-03
3	82	100.0	452	1	US-08-280- Sequence 4, Applicatio	1.05e-03
4	82	100.0	452	1	US-07-618- Sequence 2, Applicatio	1.05e-03
5	82	100.0	452	1	US-07-618- Sequence 4, Applicatio	1.05e-03
6	82	100.0	452	1	US-08-110- Sequence 8, Applicatio	1.05e-03
7	82	100.0	452	1	US-08-280- Sequence 2, Applicatio	1.05e-03
8	82	100.0	618	2	US-08-668- Sequence 5, Applicatio	1.05e-03
9	50	61.0	504	2	US-08-441- Sequence 18, Applicati	1.80e+01
10	50	61.0	521	2	US-08-557- Sequence 32, Applicati	1.80e+01
11	50	61.0	530	2	US-08-557- Sequence 35, Applicati	1.80e+01
12	50	61.0	3052	2	US-08-557- Sequence 26, Applicati	1.80e+01
13	49	59.8	1218	2	US-08-400- Sequence 6, Applicatio	2.39e+01
14	47	57.3	776	2	US-08-870- Sequence 17, Applicati	4.19e+01
15	47	57.3	776	1	US-08-198- Sequence 17, Applicati	4.19e+01
16	47	57.3	821	1	US-08-198- Sequence 6, Applicatio	4.19e+01
17	47	57.3	821	2	US-08-870- Sequence 6, Applicatio	4.19e+01
18	46	56.1	306	1	US-08-696- Sequence 4, Applicatio	5.53e+01
19	46	56.1	404	1	US-08-696- Sequence 2, Applicatio	5.53e+01
20	46	56.1	555	2	US-08-469- Sequence 27, Applicati	5.53e+01
21	45	56.1	555	2	US-08-469- Sequence 28, Applicati	5.53e+01
22	45	54.9	30	1	US-08-145- Sequence 19, Applicati	7.28e+01
23	45	54.9	3666	2	US-08-222- Sequence 12, Applicati	7.28e+01

24	45	54.9	3727	2	US-08-222- Sequence 27, Applicati	7.28e+01
25	45	54.9	3778	2	US-08-222- Sequence 2, Applicatio	7.28e+01
26	43	52.4	153	4	5196194-20 Patent No. 5196194.	1.25e-02
27	43	52.4	708	1	US-08-396- Sequence 8, Applicatio	1.25e-02
28	43	52.4	708	1	US-08-818- Sequence 8, Applicatio	1.25e-02
29	43	52.4	739	1	US-08-818- Sequence 10, Applicati	1.25e-02
30	43	52.4	739	1	US-08-396- Sequence 10, Applicati	1.25e-02
31	43	52.4	1068	1	US-08-818- Sequence 12, Applicati	1.25e-02
32	43	52.4	1068	1	US-08-396- Sequence 12, Applicati	1.25e-02
33	43	52.4	1075	3	PCT-US94-0 Sequence 41, Applicati	1.25e-02
34	42	51.2	15	4	5183734-5 Patent No. 5183734.	1.64e-02
35	42	51.2	16	4	5183734-13 Patent No. 5183734.	1.64e-02
36	42	51.2	16	4	5183734-12 Patent No. 5183734.	1.64e-02
37	42	51.2	112	2	US-08-961- Sequence 7, Applicatio	1.64e-02
38	42	51.2	112	2	US-08-283- Sequence 7, Applicatio	1.64e-02
39	42	51.2	112	2	US-08-283- Sequence 2, Applicatio	1.64e-02
40	42	51.2	112	1	US-08-326- Sequence 17, Applicati	1.64e-02
41	42	51.2	153	4	5204446-4 Patent No. 5204446.	1.64e-02
42	42	51.2	154	4	5183734-1 Patent No. 5183734.	1.64e-02
43	42	51.2	202	3	PCT-US96-1 Sequence 16, Applicati	1.64e-02
44	42	51.2	568	3	PCT-US94-0 Sequence 30, Applicati	1.64e-02
45	42	51.2	1528	1	US-08-326- Sequence 2, Applicatio	1.64e-02

ALIGNMENTS

RESULT 1
ID US-08-787-547-46 STANDARD; PRT; 12 AA.
XX
AC xxxxxx
XX
DT
XX
DE
XX
Sequence 46, Application US/08787547
Patent No. 5783567
GENERAL INFORMATION:
APPLICANT: Hedley, Mary Lynne
APPLICANT: Curley, Joanne M.
APPLICANT: Langer, Robert S.
TITLE OF INVENTION: MICROPARTICLES FOR DELIVERY
TITLE OF INVENTION: OF NUCLEIC ACID
NUMBER OF SEQUENCES: 107
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson, P.C.
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/787,547
FILING DATE: 22-JAN-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 08191/003001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-542-5070
TELEFAX: 617-542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:

CC LENGTH: 12 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
SQ SEQUENCE 12 AA; 1311 MW; 550 CN;

Query Match 100.0%; Score 82; DB 2; Length 12;
Best Local Similarity 100.0%; Pred. No. 1.05e-03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 GOIGNDPNRDIL 12
QY 1 GOIGNDPNRDIL 12

RESULT 2
ID US-08-305-871A-6 STANDARD; PRT: 13 AA.
XX
AC xxxxxx

Sequence 6, Application US/08305871A

Sequence 6, Application US/08305871A
Patent No. 5736142
GENERAL INFORMATION:
APPLICANT: Sette, Alessandro
APPLICANT: Gaeta, Federico
APPLICANT: Grey, Howard M.
APPLICANT: Sidney, John
APPLICANT: Alexander, Jeffrey L.
TITLE OF INVENTION: Alteration of Immune Response Using Pan
DR-Binding Peptides
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/305.871A
FILING DATE: 14-SEP-1994
CLASSIFICATION: 424

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/121,101
FILING DATE: 14-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Bastian, Kevin L.
REGISTRATION NUMBER: 34,774
REFERENCE/DOCKET NUMBER: 14137-0062-10
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 13 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE 13 AA; 1426 MW; 637 CN;

Query Match 100.0%; Score 82; DB 1; Length 13;
Best Local Similarity 100.0%; Pred. No. 1.05e-03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2 GOIGNDPNRDIL 13
QY 1 GOIGNDPNRDIL 12

RESULT 3
ID US-08-280-228-4 STANDARD; PRT: 452 AA.
XX
AC xxxxxx

Sequence 4, Application US/08280228

Sequence 4, Application US/08280228
Patent No. 5571694
GENERAL INFORMATION:
APPLICANT: Makoff Dr, Andrew J
APPLICANT: Romanos Dr, Michael A
APPLICANT: Clare Dr, Jeffrey J
APPLICANT: Fairweather Dr, Neil F
TITLE OF INVENTION: VACCINES
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHVE P.C.
STREET: 1100 No. 5571694th Glebe Road
CITY: Arlington,
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22201-4714

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/280.228
FILING DATE: 25-JUL-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/618,312
FILING DATE: 27-NOV-1990
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 8926832.0
FILING DATE: 28-NOV-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9006097.1
FILING DATE: 17-MAR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Wilson, Mary J.
REGISTRATION NUMBER: 32,955
REFERENCE/DOCKET NUMBER: 117-163
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
TELEX: 200797 NIXN UR
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 452 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE 452 AA; 51770 MW; 1065961 CN;

Query Match 100.0%; Score 82; DB 1; Length 452;
Best Local Similarity 100.0%; Pred. No. 1.05e-03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 410 GOIGNDPNRDIL 421
QY 1 GOIGNDPNRDIL 12

```

RESULT      4
ID   US-07-618-312A-2      STANDARD;      PRT;      452 AA.
XX
AC      xxxxxx
XX
DT
XX
DE
DE
XX
XX
Sequence 2, Application US/07618312A
CC
CC
Sequence 2, Application US/07618312A
CC
CC
Patent No. 5389540
CC
GENERAL INFORMATION:
CC
APPLICANT: Makoff Dr, Andrew J
CC
APPLICANT: Romanos Dr, Michael A
CC
APPLICANT: Clare Dr, Jeffrey J
CC
APPLICANT: Fairweather Dr, Neil F
CC
TITLE OF INVENTION: VACCINES
CC
NUMBER OF SEQUENCES: 13
CC
CORRESPONDENCE ADDRESS:
CC
ADDRESSEE: 14th Floor
CC
STREET: 2200 Clarendon Boulevard,
CC
CITY: Arlington,
CC
STATE: Virginia
CC
COUNTRY: U.S.A.
CC
ZIP: 22201
CC
COMPUTER READABLE FORM:
CC
MEDIUM TYPE: Floppy disk
CC
COMPUTER: IBM PC compatible
CC
OPERATING SYSTEM: PC-DOS/MS-DOS
CC
SOFTWARE: Patent In Release #1.0, Version #1.25
CC
CURRENT APPLICATION DATA:
CC
APPLICATION NUMBER: US/07/618,312A
CC
FILING DATE: 19910516
CC
CLASSIFICATION: 424
CC
PRIOR APPLICATION DATA:
CC
APPLICATION NUMBER: GB 8926832.0
CC
FILING DATE: 28-NOV-1989
CC
PRIOR APPLICATION DATA:
CC
APPLICATION NUMBER: GB 9006097.1
CC
FILING DATE: 17-MAR-1990
CC
ATTORNEY/AGENT INFORMATION:
CC
NAME: Crawford Mr, Arthur R
CC
REGISTRATION NUMBER: 25,327
CC
REFERENCE/DOCKET NUMBER: 510-51
CC
TELECOMMUNICATION INFORMATION:
CC
TELEPHONE: 0101 703 8750400
CC
TELEFAX: 0101 703 5253468
CC
TELEX: 200797 NIXN UR
CC
INFORMATION FOR SEQ ID NO: 2:
CC
SEQUENCE CHARACTERISTICS:
CC
LENGTH: 452 amino acids
CC
TYPE: AMINO ACID
CC
TOPOLOGY: linear
CC
MOLECULE TYPE: protein
CC
SEQUENCE 452 AA; 51797 MW; 1065526 CN;

Query Match      100.0%; Score 82; DB 1; Length 452;
Best Local Similarity 100.0%; Pred. No. 1.05e-03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 410 GQIGNDPNRDIL 421
QY 1 GQIGNDPNRDIL 12

RESULT      5
ID   US-07-618-312A-4      STANDARD;      PRT;      452 AA.
XX
AC      xxxxxx
XX
DT
XX
DE
DE
XX
XX
Sequence 8, Application US/08110786A
CC
CC
Sequence 8, Application US/08110786A
CC
CC
Patent No. 5443966
CC
GENERAL INFORMATION:
CC
APPLICANT: FAIRWEATHER, Neil Fraser
CC
APPLICANT: MAKOFF, Andrew Joseph

```

```

XX
DE
XX
Sequence 4, Application US/07618312A
CC
CC
Sequence 4, Application US/07618312A
CC
CC
Patent No. 5389540
CC
GENERAL INFORMATION:
CC
APPLICANT: Makoff Dr, Andrew J
CC
APPLICANT: Romanos Dr, Michael A
CC
APPLICANT: Clare Dr, Jeffrey J
CC
APPLICANT: Fairweather Dr, Neil F
CC
TITLE OF INVENTION: VACCINES
CC
NUMBER OF SEQUENCES: 13
CC
CORRESPONDENCE ADDRESS:
CC
ADDRESSEE: 14th Floor
CC
STREET: 2200 Clarendon Boulevard,
CC
CITY: Arlington,
CC
STATE: Virginia
CC
COUNTRY: U.S.A.
CC
ZIP: 22201
CC
COMPUTER READABLE FORM:
CC
MEDIUM TYPE: Floppy disk
CC
COMPUTER: IBM PC compatible
CC
OPERATING SYSTEM: PC-DOS/MS-DOS
CC
SOFTWARE: Patent In Release #1.0, Version #1.25
CC
CURRENT APPLICATION DATA:
CC
APPLICATION NUMBER: US/07/618,312A
CC
FILING DATE: 19910516
CC
CLASSIFICATION: 424
CC
PRIOR APPLICATION DATA:
CC
APPLICATION NUMBER: GB 8926832.0
CC
FILING DATE: 28-NOV-1989
CC
PRIOR APPLICATION DATA:
CC
APPLICATION NUMBER: GB 9006097.1
CC
FILING DATE: 17-MAR-1990
CC
ATTORNEY/AGENT INFORMATION:
CC
NAME: Crawford Mr, Arthur R
CC
REGISTRATION NUMBER: 25,327
CC
REFERENCE/DOCKET NUMBER: 510-51
CC
TELECOMMUNICATION INFORMATION:
CC
TELEPHONE: 0101 703 8750400
CC
TELEFAX: 0101 703 5253468
CC
TELEX: 200797 NIXN UR
CC
INFORMATION FOR SEQ ID NO: 4:
CC
SEQUENCE CHARACTERISTICS:
CC
LENGTH: 452 amino acids
CC
TYPE: AMINO ACID
CC
TOPOLOGY: linear
CC
MOLECULE TYPE: protein
CC
SEQUENCE 452 AA; 51770 MW; 1065961 CN;

Query Match      100.0%; Score 82; DB 1; Length 452;
Best Local Similarity 100.0%; Pred. No. 1.05e-03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 410 GQIGNDPNRDIL 421
QY 1 GQIGNDPNRDIL 12

RESULT      6
ID   US-08-110-786A-8      STANDARD;      PRT;      452 AA.
XX
AC      xxxxxx
XX
DT
XX
XX
Sequence 8, Application US/08110786A
CC
CC
Sequence 8, Application US/08110786A
CC
CC
Patent No. 5443966
CC
GENERAL INFORMATION:
CC
APPLICANT: FAIRWEATHER, Neil Fraser
CC
APPLICANT: MAKOFF, Andrew Joseph

```

CC TITLE OF INVENTION: Expression of tetanus toxin fragment C
CC NUMBER OF SEQUENCES: 8
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Nixon & Vanderhye P.C.
CC STREET: 1100 No. 5443966th Glebe Road
CC CITY: Arlington
CC STATE: Virginia
CC COUNTRY: U.S.A.
CC ZIP: 22201-4714
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/110,786A
CC FILING DATE: 23-AUG-1993 1991
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/777,337
CC FILING DATE: 29-NOV-1991
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: PCT/GB90/00943
CC FILING DATE: 20-JUN-1990
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: GB 8914122.0
CC FILING DATE: 20 June 1989
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Mary J. Wilson
CC REGISTRATION NUMBER: 32,955
CC REFERENCE/DOCKET NUMBER: 117-134
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (703) 816-4000
CC TELEFAX: (703) 816-4100
CC TELEX: 200797 NIXN UR
CC INFORMATION FOR SEQ ID NO: 8:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 452 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 452 AA; 51784 MW; 1060967 CN;
Query Match 100.0%; Score 82; DB 1; Length 452;
Best Local Similarity 100.0%; Pred. No. 1.05e-03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 410 GQIGNDPNRDIL 421
QY 1 GQIGNDPNRDIL 12
RESULT 7
ID US-08-280-228-2 STANDARD; PRT: 452 AA.
XX xxxxxx
XX
XX
XX
XX
DE Sequence 2, Application US/08280228
DE Sequence 2, Application US/08280228
XX Sequence 2, Application US/08280228
CC Patent No. 5571694
CC GENERAL INFORMATION:
CC APPLICANT: Makoff Dr, Andrew J
CC APPLICANT: Romanos Dr, Michael A
CC APPLICANT: Clare Dr, Jeffrey J
CC APPLICANT: Fairweather Dr, Neil F
CC TITLE OF INVENTION: VACCINES
CC NUMBER OF SEQUENCES: 13
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: NIXON & VANDERHYE P.C.
CC STREET: 1100 No. 5571694th Glebe Road

CC CITY: Arlington,
CC STATE: Virginia
CC COUNTRY: U.S.A.
CC ZIP: 22201-4714
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/280,228
CC FILING DATE: 25-JUL-1994
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/618,312
CC FILING DATE: 27-NOV-1990
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: GB 8926832.0
CC FILING DATE: 28-NOV-1989
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: GB 9006097.1
CC FILING DATE: 17-MAR-1990
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Wilson, Mary J.
CC REGISTRATION NUMBER: 32,955
CC REFERENCE/DOCKET NUMBER: 117-163
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (703) 816-4000
CC TELEFAX: (703) 816-4100
CC TELEX: 200797 NIXN UR
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 452 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 452 AA; 51797 MW; 1065526 CN;
Query Match 100.0%; Score 82; DB 1; Length 452;
Best Local Similarity 100.0%; Pred. No. 1.05e-03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 410 GQIGNDPNRDIL 421
QY 1 GQIGNDPNRDIL 12
RESULT 8
ID US-08-668-381A-5 STANDARD; PRT: 618 AA.
XX xxxxxx
XX
XX
XX
XX
XX
DE Sequence 5, Application US/08668381A
DE Sequence 5, Application US/08668381A
XX Patent No. 5780024
XX GENERAL INFORMATION:
XX APPLICANT: Brown, Robert H.
XX APPLICANT: Fishman, Paul S.
XX APPLICANT: Francis, Jonathan W.
XX APPLICANT: Hosler, Betsy A.
XX TITLE OF INVENTION: SUPEROXIDE DISMUTASE/TETANUS TOXIN
XX TITLE OF INVENTION: FRAGMENT C HYBRID PROTEIN
XX NUMBER OF SEQUENCES: 6
XX CORRESPONDENCE ADDRESS:
XX ADDRESSEE: Fish & Richardson P.C.
XX STREET: 225 Franklin Street
XX CITY: Boston
XX STATE: MA
XX COUNTRY: USA

CC ZIP: 02110-2804
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC OPERATING SYSTEM: IBM PC compatible
CC SOFTWARE: Patent in Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/668,381A
CC FILING DATE: 21-JUN-1996
CC CLASSIFICATION: 514
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 60/000,473
CC FILING DATE: 23-JUN-1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Clark, Paul T.
CC REGISTRATION NUMBER: 30,164
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 617/542-5070
CC TELEFAX: 617/542-8906
CC TELEX: 200154
CC INFORMATION FOR SEQ ID NO: 5:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 618 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 618 AA; 68895 MW; 1991829 CN;
Query Match 100.0%; Score 82; DB 2; Length 618;
Best Local Similarity 100.0%; Pred. No. 1.05e-03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 576 GQICNDPNDIL 587
QY 1 GQICNDPNDIL 12
RESULT 9
ID US-08-441-139-18 STANDARD; PRT; 504 AA.
XX
AC xxxxxx
XX
DT
DE Sequence 18, Application US/08441139
XX Sequence 18, Application US/08441139
XX Patent No. 5773245
CC GENERAL INFORMATION:
CC APPLICANT: Wittrup, Dr. Karl D.
CC APPLICANT: Robinson, Anne S.
CC TITLE OF INVENTION: METHODS FOR INCREASING SECRETION OF
CC TITLE OF INVENTION: RECOMBINANTLY EXPRESSED PROTEINS
CC NUMBER OF SEQUENCES: 20
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
CC STREET: 400 Garden City Plaza
CC CITY: Garden City
CC STATE: NY
CC COUNTRY: USA
CC ZIP: 11530
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent in Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/441,139
CC FILING DATE: 15-MAY-1995
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/089,997

CC FILING DATE: 06-JUL-1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Digiglio, Frank S.
CC REGISTRATION NUMBER: 31,346
CC REFERENCE/DOCKET NUMBER: 8646
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 516-742-4343
CC TELEFAX: 516-742-4366
CC TELEX: 230 901 SANS UR
CC INFORMATION FOR SEQ ID NO: 18:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 504 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 504 AA; 55937 MW; 1256202 CN;
Query Match 61.0%; Score 50; DB 2; Length 504;
Best Local Similarity 54.5%; Pred. No. 1.80e-01;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
Db 371 EIVNDPKDVL 381
QY 2 QICNDPNDIL 12
RESULT 10
ID US-08-557-122A-32 STANDARD; PRT; 521 AA.
XX
AC xxxxxx
XX
DT
DE Sequence 32, Application US/08557122A
XX Sequence 32, Application US/08557122A
XX Patent No. 5879664
CC GENERAL INFORMATION:
CC APPLICANT: Hjort, Carsten Mailand
CC TITLE OF INVENTION: Fungal Protein Disulfide Isomerase
CC NUMBER OF SEQUENCES: 38
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: No. 5879664o No. 5879664disk of No. 5879664th America, Inc.
CC STREET: 405 Lexington Avenue, 64th Floor
CC CITY: New York
CC STATE: New York
CC COUNTRY: United States of America
CC ZIP: 10174-6401
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent in Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/557,122A
CC FILING DATE: 11-DEC-1995
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Lambiris, Elias J.
CC REGISTRATION NUMBER: 33,728
CC REFERENCE/DOCKET NUMBER: 3980.204-US
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 212-867-0123
CC TELEFAX: 212-878-9655
CC INFORMATION FOR SEQ ID NO: 32:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 521 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 521 AA; 58129 MW; 1344734 CN;

Query Match 61.0%; Score 50; DB 2; Length 521;
Best Local Similarity 54.5%; Pred. No. 1.80e+01;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 387 EIVNDPKKDV 397
:||||:|
QY 2 QIGNDPNRDIL 12

RESULT 11
ID US-08-557-122A-35 STANDARD; PRT; 530 AA.
XX
AC xxxxxx
XX
DI
XX
DE

Sequence 35, Application US/08557122A

Sequence 35, Application US/08557122A

Patent No. 5879664

GENERAL INFORMATION:

APPLICANT: Hjort, Carsten Mailand

TITLE OF INVENTION: Fungal Protein Disulfide Isomerase

NUMBER OF SEQUENCES: 38

CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 5879664 No. 5879664disk of No. 5879664th America, Inc.

STREET: 405 Lexington Avenue, 64th Floor

CITY: New York

STATE: New York

COUNTRY: United States of America

ZIP: 10174-6401

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/557,122A

FILING DATE: 11-DEC-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Lambiris, Elias J.

REGISTRATION NUMBER: 33,728

REFERENCE/DOCKET NUMBER: 3980.204-US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-867-0123

TELEFAX: 212-878-9655

INFORMATION FOR SEQ ID NO: 35:

SEQUENCE CHARACTERISTICS:

LENGTH: 530 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: Peptide

SEQUENCE 530 AA; 59082 MW; 1369542 CN;

Query Match 61.0%; Score 50; DB 2; Length 530;
Best Local Similarity 54.5%; Pred. No. 1.80e+01;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 388 EIVNDPKKDV 398
:||||:|
QY 2 QIGNDPNRDIL 12

RESULT 12
ID US-08-557-122A-26 STANDARD; PRT; 3052 AA.
XX
AC xxxxxx
XX
XX
DT
XX
DE

Sequence 26, Application US/08557122A

XX Sequence 26, Application US/08557122A
CC Patent No. 5879664
CC GENERAL INFORMATION:
CC APPLICANT: Hjort, Carsten Mailand
CC TITLE OF INVENTION: Fungal Protein Disulfide Isomerase
CC NUMBER OF SEQUENCES: 38
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: No. 5879664 No. 5879664disk of No. 5879664th America, Inc.
CC STREET: 405 Lexington Avenue, 64th Floor
CC CITY: New York
CC STATE: New York
CC COUNTRY: United States of America
CC ZIP: 10174-6401
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/557,122A
CC FILING DATE: 11-DEC-1995
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Lambiris, Elias J.
CC REGISTRATION NUMBER: 33,728
CC REFERENCE/DOCKET NUMBER: 3980.204-US
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 212-867-0123
CC TELEFAX: 212-878-9655
CC INFORMATION FOR SEQ ID NO: 26:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 3052 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: Single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 3052 AA; 341596 MW; 45880055 CN;

Query Match 61.0%; Score 50; DB 2; Length 3052;
Best Local Similarity 54.5%; Pred. No. 1.80e+01;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 2497 EIVNDPKKDV 2507
:||||:|
QY 2 QIGNDPNRDIL 12

RESULT 13
ID US-08-400-159-6 STANDARD; PRT; 1218 AA.
XX
AC xxxxxx
XX
DT
XX
DE

Sequence 6, Application US/08400159

Sequence 6, Application US/08400159

Patent No. 5869282

GENERAL INFORMATION:

APPLICANT: Ish-Horowitz, David

APPLICANT: Henrique, Domingos M.P.

APPLICANT: Lewis, Julian H.

APPLICANT: Myat, Anna M.

APPLICANT: Fleming, Robert J.

APPLICANT: Artavanis-Tsakonas, Spyridon

APPLICANT: Mann, Robert S.

APPLICANT: Gray, Grace E.

TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF THE

TITLE OF INVENTION: SERRATE GENE AND METHODS BASED THEREON

NUMBER OF SEQUENCES: 20

CORRESPONDENCE ADDRESS:

ADDRESSEE: pennie & Edmonds

CC STREET: 1155 Avenue of the Americas
CC CITY: New York
CC STATE: New York
CC COUNTRY: USA
CC ZIP: 10036-2711
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/400,159
CC FILING DATE: 07-MAR-1995
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Misrock, S. Leslie
CC REGISTRATION NUMBER: 18,872
CC REFERENCE/DOCKET NUMBER: 7326-029
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 790-9090
CC TELEFAX: (212) 869-9741/8864
CC TELEX: 66141 PENNIE
CC INFORMATION FOR SEQ ID NO: 6:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 1218 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC QUERY MATCH 59.8%; Score 49; DB 2; Length 1218;
CC Best Local Similarity 77.8%; Pred. No. 2.39e+01;
CC Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
SQ SEQUENCE 1218 AA; 133810 MW; 7064054 CN;

Db 114 GNDPNRIVL 122
YQ 4 GNDPNRDIL 12

RESULT 14
ID US-08-870-593-17 STANDARD; PRT: 776 AA.
XX
AC xxxxxx
DE DE
DT DT
XX
XX
Sequence 17, Application US/08870693
Sequence 17, Application US/08870593
Patent No. 5866338
GENERAL INFORMATION:
APPLICANT: Hartwell, Leland H.
APPLICANT: Weinert, Ted A.
APPLICANT: Pilon, Sharon E.
APPLICANT: Groudine, Mark T.
TITLE OF INVENTION: Cell Cycle Checkpoint Genes
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC
STREET: 1420 Fifth Ave., Suite 2800
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101-2347
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/870,693
FILING DATE:
CLASSIFICATION: 536

CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/198,446
CC FILING DATE: February 18, 1994
CC APPLICATION NUMBER: PCT/US93/04458
CC FILING DATE: May 12, 1993
CC APPLICATION NUMBER: US 07/884,426
CC FILING DATE: May 14, 1992
CC APPLICATION NUMBER: US 07/882,051
CC FILING DATE: May 12, 1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Sheiness, Diana K.
CC REGISTRATION NUMBER: 35,356
CC REFERENCE/DOCKET NUMBER: FHCRI10798
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 206-682-8100
CC TELEFAX: 206-224-0779
CC INFORMATION FOR SEQ ID NO: 17:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 776 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC DESCRIPTION: yeast MEC2 protein
CC QUERY MATCH 57.3%; Score 47; DB 2; Length 776;
CC Best Local Similarity 45.5%; Pred. No. 4.19e+01;
CC Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
SQ SEQUENCE 776 AA; 87015 MW; 3041193 CN;

Db 289 GAVGEDAGREI 299
YQ 1 GQIGNDPNRDI 11

RESULT 15
ID US-08-198-446B-17 STANDARD; PRT: 776 AA.
XX
AC xxxxxx
DE DE
DT DT
XX
XX
Sequence 17, Application US/08198446B
Sequence 17, Application US/08198446B
Patent No. 5674996
GENERAL INFORMATION:
APPLICANT: Hartwell, Leland H.
APPLICANT: Weinert, Ted A.
APPLICANT: Pilon, Sharon E.
APPLICANT: Groudine, Mark T.
TITLE OF INVENTION: Cell Cycle Checkpoint Genes
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC
STREET: 1420 Fifth Ave., Suite 2800
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101-2347
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/198,446B
FILING DATE: 18-FEB-1994
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Sheiness, Diana K.
REGISTRATION NUMBER: 35,356
REFERENCE/DOCKET NUMBER: FHCRI17537
TELECOMMUNICATION INFORMATION:

CC TELEPHONE: 206-682-8100
CC TELEFAX: 206-224-0779
CC INFORMATION FOR SEQ ID NO: 17:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 776 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC DESCRIPTION: yeast MEC2 protein
SQ SEQUENCE 776 AA; 87015 MW; 3041193 CN;

Query Match 57.3%; Score 47; DB 1; Length 776;
Best Local Similarity 45.5%; Pred. No. 4.19e+01;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 289 GAVGEDAGREI 299
QY 1 GQIGNDPNRDI 11

Search completed: Tue Aug 17 16:00:34 1999
Job time : 7 secs.

M I S R E L H

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Aug 17 16:02:46 1999; MasPar time 4.35 Seconds

Tabular output not generated. 119.724 Million cell updates/sec

Title: >US-09-049-847-4
Description: (1-13) from US09049847.pep
Perfect Score: 104
Sequence: 1 KFLAVWKITYKDT 13

Scoring table: PAM 150
Gap 15

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 26.787; Variance 42.449; scale 0.631

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	87	83.7	2207	1	genome polyprotein (v	7.13e-05
2	87	83.7	2209	1	genome polyprotein (v	7.13e-05
3	85	81.7	2205	1	genome polyprotein -	1.75e-04
4	85	81.7	2207	1	genome polyprotein -	1.75e-04
5	85	81.7	2209	1	genome polyprotein -	1.75e-04
6	82	78.8	2207	2	genome polyprotein -	6.68e-04
7	79	76.0	2206	1	genome polyprotein -	2.49e-03
8	79	76.0	2206	2	genome polyprotein -	2.49e-03
9	77	74.0	2206	1	genome polyprotein -	5.92e-03
10	70	67.3	1040	2	receptor tyrosine kin	1.13e-01
11	67	64.4	1040	2	conserved hypothetical	3.82e-01
12	64	61.5	2206	1	genome polyprotein -	1.25e+00
13	62	59.6	874	2	DNA polymerase III al	2.73e+00
14	61	58.7	149	2	hemoglobin - southern	4.00e+00
15	61	58.7	255	2	hypothetical protein	4.00e+00
16	59	56.7	940	2	hypothetical protein	4.00e+00
17	59	56.7	345	2	probable lipase - syn	8.48e+00
18	59	56.7	741	2	myosin heavy chain, e	8.48e+00
19	59	56.7	858	2	myosin heavy chain, a	8.48e+00
20	59	56.7	955	2	myosin heavy chain, e	8.48e+00
21	59	56.7	1938	1	myosin heavy chain, f	8.48e+00
22	58	55.8	22	2	light meromyosin - ch	1.23e+01
23	58	55.8	847	2	alanine--trna ligase	1.23e+01

ALIGNMENTS

RESULT 1
ENTRY
TITLE
CONTAINS
GNNYLP #type complete
genome polyprotein (version 1) - human poliovirus 1 (strain Mahoney)
coat protein VP1; coat protein VP2; coat protein VP3; coat protein VP4; core protein P2-3b; core protein P2-5b; core protein P2-X; genome-linked protein VPg; Probable RNA polymerase P3-7c; protein P3-1b; protein P3-2; RNA-directed RNA polymerase (EC 2.7.7.48) P3-4b
ORGANISM #formal_name human poliovirus 1
DATE 17-Dec-1982 #sequence_revision 17-Dec-1982 #text_change 05-Jun-1998
ACCESSIONS A93258; A90800; A03897
REFERENCE A93258
#authors Kitamura, N.; Semler, B.L.; Rothberg, P.G.; Larsen, G.R.; Adler, C.J.; Dorner, A.J.; Emami, E.A.; Hanecak, R.; Lee, J.J.; van der Werf, S.; Anderson, C.W.; Wimmer, E.
#journal Nature (1981) 291:547-553
#title Primary structure, gene organization and polypeptide expression of poliovirus RNA.
#cross-references MUID:81220953
#accession A93258
#molecule_type genomic RNA
#residues 1-2207 #label KIT1
#cross-references GB:V01148; NID:961236; PID:961237
#note the amino acid sequence of VPg (residues 1543-1564) was also determined and agrees with that shown

REFERENCE A90800
#authors Kitamura, N.; Adler, C.J.; Rothberg, P.G.; Martinko, J.; Nathanson, S.G.; Wimmer, E.
#journal Cell (1980) 21:295-302
#title The genome-linked protein of picornaviruses. VII. Genetic mapping of poliovirus VPg by protein and RNA sequence studies.
#cross-references MUID:81001866
#accession A90800
#molecule_type genomic RNA
#residues 1539-1574 #label KIT2
#note the amino end of VPg corresponds to residue 1543; a choice between the two potential carboxyl cleavage sites, after residue 1564 or 1569, could not be made; the partial sequence of this protein obtained by radiochemical microsequence analysis agrees with that predicted by the virion RNA
REFERENCE A30637
#authors Rothberg, P.G.; Harris, T.J.; Nomoto, A.; Wimmer, E.
#journal Proc. Natl. Acad. Sci. U.S.A. (1978) 75:4868-4872

```

#title
#contents
COMMENT VPg is linked by Tyr-1545 to the uridylylate residue at the 5' end of
the genome RNA. It is required to initiate RNA synthesis and it
may also be involved in morphogenesis.
COMMENT Coat proteins VP2 and VP3 and the RNA-directed RNA polymerase are
related to their counterparts in foot-and-mouth disease virus.
CLASSIFICATION #superfamily poliovirus genome polyprotein
KEYWORDS genome-linked protein; nucleotidyltransferase;
phosphoprotein; polyprotein
FEATURE
2-69 #product coat protein VP4 #status predicted #label VP4\
70-340 #product coat protein VP2 #status predicted #label VP2\
341-578 #product coat protein VP3 #status predicted #label VP3\
579-880 #product coat protein VP1 #status predicted #label VP1\
881-1455 #product core protein P2-3b #status predicted #label
P23\
1030-1455 #product core protein P2-5b #status predicted #label
P25\
1127-1455 #product core protein P2-X #status predicted #label P2X\
1456-2207 #product protein P3-1b #status predicted #label P31\
1543-1564 #product genome-linked protein VPg #status predicted
#label VPg\
1555-2207 #product protein P3-2 #status predicted #label P32\
1565-1746 #product probable proteinase P3-7c #status predicted
#label P37\
1747-2207 #product RNA-directed RNA polymerase P3-4b #status
#predicted #label P34\
1545 #binding_site phosphoryl-RNA (Tyr) (covalent) #status
experimental
SUMMARY #length 2207 #molecular-weight 246540 #checksum 7480
Query Match 83.7%; Score 87; DB 1; Length 2207;
Best Local Similarity 84.6%; Pred. No. 7.13e-05;
Matches 11; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 681 KLFVAVKITYKDT 693
QY 1 KLFVAVKITYKDT 13
|:::|||||
RESULT 2
ENTRY #type complete
TITLE genome polyprotein (version 2) - human poliovirus 1 (strain
Mahoney)
ORGANISM #formal_name human poliovirus 1
DATE 18-Dec-1981 #sequence_revision 18-Dec-1981 #text_change
05-Jun-1998
ACCESSION A03898
REFERENCE #authors Racanelli, V.R.; Baltimore, D.
#journal Proc. Natl. Acad. Sci. U.S.A. (1981) 78:4887-4891
#title Molecular cloning of poliovirus cDNA and determination of the
complete nucleotide sequence of the viral genome.
#cross-references MUID:82050159
#accession A03898
#molecule_type genomic RNA
#residues 1-2209 #label RAC
#cross-references GB:J02281; NID:g332886; PID:g332887
#note the authors translated the codon TTC for residue 464 as
Ser
REFERENCE A0636
#authors Ambros, V.; Baltimore, D.
#journal J. Biol. Chem. (1978) 253:5263-5266
#title Protein is linked to the 5' end of poliovirus RNA by a
phosphodiester linkage to tyrosine.
#contents annotation; chemical characterization
CLASSIFICATION #superfamily poliovirus genome polyprotein
KEYWORDS genome-linked protein; nucleotidyltransferase;
phosphoprotein; polyprotein
FEATURE
2-69 #product coat protein VP4 #status predicted #label VP4\

```

```

70-341 #product coat protein VP2 #status predicted #label VP2\
342-579 #product coat protein VP3 #status predicted #label VP3\
580-881 #product coat protein VP1 #status predicted #label VP1\
882-1456 #product core protein P2-3b #status predicted #label
P23\
1031-1456 #product core protein P2-5b #status predicted #label
P25\
1128-1456 #product core protein P2-X #status predicted #label P2X\
1457-2209 #product protein P3-1b #status predicted #label P31\
1544-1565 #product genome-linked protein VPg #status predicted
#label VPg\
1566-2209 #product protein P3-2 #status predicted #label P32\
1566-1748 #product probable proteinase P3-7c #status predicted
#label P37\
1749-2209 #product RNA-directed RNA polymerase P3-4b #status
#predicted #label P34\
1546 #binding_site phosphoryl-RNA (Tyr) (covalent) #status
experimental
SUMMARY #length 2209 #molecular-weight 246538 #checksum 7583
Query Match 83.7%; Score 87; DB 1; Length 2209;
Best Local Similarity 84.6%; Pred. No. 7.13e-05;
Matches 11; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 682 KLFVAVKITYKDT 694
QY 1 KLFVAVKITYKDT 13
|:::|||||
RESULT 3
ENTRY #type complete
TITLE genome polyprotein - human poliovirus 2 (strain W-2)
CONTAINS coat protein 1A; coat protein 1B; coat protein 1C; coat
protein 1D; genome-linked protein VPg; nonstructural
protein 2B; nonstructural protein 2C; nonstructural protein
3A; proteinase 2A; proteinase 3C; RNA-directed RNA
polymerase (EC 2.7.7.48)
ORGANISM #formal_name human poliovirus 2
DATE 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change
05-Jun-1998
ACCESSION A34032
REFERENCE #authors Pevear, D.C.; Oh, C.K.; Cunningham, L.L.; Calenoff, M.;
Jubelt, B.
#journal J. Gen. Virol. (1990) 71:43-52
#title Localization of genomic regions specific for the attenuated,
mouse-adapted poliovirus type 2 strain W-2.
#cross-references MUID:90155230
#accession A34032
#molecule_type genomic RNA
#residues 1-2205 #label PEV
CLASSIFICATION #superfamily poliovirus genome polyprotein
KEYWORDS coat protein; genome-linked protein; nonstructural protein;
nucleotidyltransferase; phosphoprotein; polyprotein;
proteinase
FEATURE
1-69 #product coat protein 1A #status predicted #label VP4\
70-340 #product coat protein 1B #status predicted #label VP2\
341-578 #product coat protein 1C #status predicted #label VP3\
579-879 #product coat protein 1D #status predicted #label VP1\
880-1028 #product proteinase 2A #status predicted #label P2A\
1029-1125 #product nonstructural protein 2B #status predicted
#label P2B\
1126-1454 #product nonstructural protein 2C #status predicted
#label P2C\
1455-1541 #product nonstructural protein 3A #status predicted
#label P3A\
1542-1563 #product genome-linked protein VPg #status predicted
#label VPg\
1564-1746 #product proteinase 3C #status predicted #label P3C\
1747-2205 #product RNA-directed RNA polymerase #status predicted
#label RRP\
#binding_site phosphoryl-RNA (Tyr) (covalent) #status
1544

```

```

SUMMARY      #length 2205 #molecular-weight 245701 #checksum 6218
              Predicted
Query Match   81.7%; Score 85; DB 1; Length 2205;
Best Local Similarity 76.9%; Pred. No. 1.75e-04;
Matches 10; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 681 KLFVWVKITYKDT 693
      ::::::::::::::
QY 1 KFLAVWKITYKDT 13

RESULT 4
ENTRY  genome polypeptide - human poliovirus 2 (strain Lansing)
TITLE   coat protein VP1; coat protein VP2; coat protein VP3; coat
CONTAINS protein VP4; genome-linked protein VPg; nonstructural
          protein 2B; nonstructural protein 2C; nonstructural protein
          3A; proteinase (EC 3.4.22.1) 2A; proteinase (EC 3.4.22.1) 3C;
          RNA-directed RNA polymerase 2
ORGANISM #formal_name human poliovirus 2
DATE      31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change
          05-Jun-1998
ACCESSION A29507
REFERENCE La Monica, N.; Meriam, C.; Racaniello, V.R.
#journal J. Virol. (1986) 57:515-525
#title Mapping of sequences required for mouse neurovirulence of
          poliovirus type 2 Lansing.
#cross-references MUID:86115399
#accession A29507
#molecule_type genomic RNA
#residues 1-2207 #label LAM
#cross-references GB:M12137; MID:g332890; PID:g332891
CLASSIFICATION #superfamily poliovirus genome polypeptide
KEYWORDS coat protein; genome-linked protein; hydrolase; nonstructural
          protein; nucleotidyltransferase; phosphoprotein;
          polypeptide; proteinase

FEATURE
1-69      #product coat protein VP4 #status predicted #label VP4\
70-340    #product coat protein VP2 #status predicted #label VP2\
341-578   #product coat protein VP3 #status predicted #label VP3\
579-879   #product coat protein VP1 #status predicted #label VP1\
880-1028  #product proteinase 2A #status predicted #label P2A\
1029-1125 #product nonstructural protein 2B #status predicted
          #label N2B\
1126-1454 #product nonstructural protein 2C #status predicted
          #label N2C\
1455-1541 #product nonstructural protein 3A #status predicted
          #label N3A\
1542-1563 #product genome-linked protein VPg #status predicted
          #label VPg\
1564-1746 #product proteinase 3C #status predicted #label P3C\
1747-2207 #product RNA-directed RNA polymerase #status predicted
          #label RPP\
1544      #binding_site phosphoryl-RNA (Tyr) (covalent) #status
          predicted

SUMMARY      #length 2207 #molecular-weight 245829 #checksum 666
              Predicted
Query Match   81.7%; Score 85; DB 1; Length 2207;
Best Local Similarity 76.9%; Pred. No. 1.75e-04;
Matches 10; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 681 KLFVWVKITYKDT 693
      ::::::::::::::
QY 1 KFLAVWKITYKDT 13

RESULT 5
ENTRY  genome polypeptide - human poliovirus 1 (strain Sabin)
TITLE   genome polypeptide - human poliovirus 1 (strain Sabin)
CONTAINS #formal_name human poliovirus 1
ORGANISM #formal_name human poliovirus 1
DATE      17-Dec-1982 #sequence_revision 17-Dec-1982 #text_change
          05-Jun-1998

```

```

22-Jan-1999
ACCESSIONS A03899
REFERENCE   A03899
#authors   Nomoto, A.; Omata, T.; Toyoda, H.; Kuge, S.; Horie, H.;
           Kataoka, Y.; Genba, Y.; Nakano, Y.; Imura, N.
#journal   Proc. Natl. Acad. Sci. U.S.A. (1982) 79:5793-5797
#title     Complete nucleotide sequence of the attenuated poliovirus
           Sabin 1 strain genome.
#cross-references MUID:83299876
#accession A03899
#molecule_type genomic RNA
#residues 1-2209 #label NOM
#cross-references GB:V01150; GB:J02282; GB:J02285; GB:J02286;
           GB:V01133; MID:g61257; PID:g61258
#note      this virus is a live vaccine strain derived from the
           Mahoney strain by spontaneous mutations during the
           attenuation process
CLASSIFICATION #superfamily poliovirus genome polypeptide
KEYWORDS genome-linked protein; phosphoprotein; polypeptide
FEATURE
2-69      #product coat protein VP4 #status predicted #label VP4\
70-341    #product coat protein VP2 #status predicted #label VP2\
342-579   #product coat protein VP3 #status predicted #label VP3\
580-881   #product coat protein VP1 #status predicted #label VP1\
882-1456  #product core protein P2-3b #status predicted #label
          P23\
1031-1456 #product core protein P2-5b #status predicted #label
          P25\
1128-1456 #product core protein P2-X #status predicted #label P2X\
1457-2209 #product protein P3-1b #status predicted #label P31\
1544-1565 #product genome-linked protein VPg #status predicted
          #label VPg\
1566-2209 #product protein P3-2 #status predicted #label P32\
1566-1748 #product probable proteinase P3-7c #status predicted
          #label P37\
1749-2209 #product RNA-directed RNA polymerase P3-4b #status
          predicted #label P34\
1546      #binding_site phosphoryl-RNA (Tyr) (covalent) #status
          predicted

SUMMARY      #length 2209 #molecular-weight 246576 #checksum 5033
              Predicted
Query Match   81.7%; Score 85; DB 1; Length 2209;
Best Local Similarity 76.9%; Pred. No. 1.75e-04;
Matches 10; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 682 KLFVWVKITYKDT 694
      ::::::::::::::
QY 1 KFLAVWKITYKDT 13

RESULT 6
ENTRY  genome polypeptide - human poliovirus 2 (strain P712)
TITLE   coat protein VP1; coat protein VP2; coat protein VP3; coat
CONTAINS protein VP4; core protein P2-3b; core protein P2-5b; core
          protein P2-X; genome-linked protein VPg; protein P3-1b;
          protein P3-2; protein P3-4a; protein P3-6a; protein P3-6b;
          protein P3-9; proteinase P3-7c; RNA-directed RNA polymerase
          (EC 2.7.7.48) P3-4b
          #formal_name human poliovirus 2
          #formal_name human poliovirus 2
          07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change
          05-Jun-1998
ACCESSIONS S09553
REFERENCE   S03822
#authors   Toyoda, H.; Kohara, M.; Kataoka, Y.; Suganuma, T.; Omata, T.;
           Imura, N.; Nomoto, A.
#journal   J. Mol. Biol. (1984) 174:561-585
#title     Complete nucleotide sequences of all three poliovirus
           serotype genomes. Implication for genetic relationship,
           gene function and antigenic determinants.
#cross-references MUID:84216300
#accession S09553
#molecule_type genomic RNA

```

```

##residues 1-2207 #label TOY
##cross-references EMBL:X00595; NID:g61127; PID:g275415; PID:g1628430
CLASSIFICATION #superfamily poliovirus genome polypeptide
KEYWORDS genome-linked protein; nucleotidyltransferase; phosphoprotein; polypeptide
FEATURE
1544 #binding_site phosphoryl-RNA (Tyr) (covalent) #status
SUMMARY #length 2207 #molecular-weight 245967 #checksum 1659
Query Match 78.8%; Score 82; DB 2; Length 2207;
Best Local Similarity 69.2%; Pred. No. 6.68e-04;
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
Db 681 RLFSVWKIIVKDT 693
QY 1 KFLAVWKIIVKDT 13
RESULT 7
ENTRY #type complete
TITLE genome polypeptide - human poliovirus 3 (strain Sabin vaccine
CONTAINS coat protein VP1; coat protein VP2; coat protein VP3; coat
protein VP4; core protein P2-3b; core protein P2-5b; core
protein P2-X; genome-linked protein VPg; probable
proteinase P3-7c; protein P3-1b; protein P3-2; RNA-directed
RNA polymerase (EC 2.7.7.48) P3-4b
ORGANISM #formal_name human poliovirus 3
DATE 19-Feb-1984 #sequence_revision 03-Aug-1984 #text_change
ACCESSION A93987; A93484; S42524; A03900
REFERENCE A93987
#authors Stanway, G.; Hughes, P.J.; Mountford, R.C.; Reeve, P.; Minor,
P.D.; Schild, G.C.; Almond, J.W.
#journal Proc. Natl. Acad. Sci. U.S.A. (1984) 81:1539-1543
#title Comparison of the complete nucleotide sequences of the
genomes of the neurovirulent poliovirus P3/Leon/37 and its
attenuated Sabin vaccine derivative P3/Leon 12a-1b.
#cross-references MUID:84170338
#accession A93987
##molecule_type genomic RNA
##residues 1-2206 #label ST1
##cross-references GB:K01392; NID:g332895; PID:g332896
##experimental_source strain Sabin vaccine P3/Leon/37
##note the strain Sabin vaccine P3/Leon/37 is the progenitor of
the strain Sabin vaccine P3/Leon 12a[1]b
##note the authors translated the codon GAU for residue 497 as
Gly
REFERENCE A93484
#authors Stanway, G.; Cann, A.J.; Hauptmann, R.; Hughes, P.; Clarke,
L.D.; Mountford, R.C.; Minor, P.D.; Schild, G.C.; Almond,
J.W.
#journal Nucleic Acids Res. (1993) 11:5629-5643
#title The nucleotide sequence of poliovirus type 3 lon 12 a-1b:
comparison with poliovirus type 1.
#cross-references MUID:83299239
#accession A93484
##molecule_type genomic RNA
##residues 1-430, F', 432-863, R', 865-907, A', 909-2206 #label ST2
##experimental_source strain Sabin vaccine P3/Leon 12a[1]b
##note the authors translated the codon GAU for residue 497 as
Gly
REFERENCE S42524
#authors Stanway, G.; Cann, A.J.; Hauptmann, R.; Mountford, R.C.;
Clarke, L.D.; Reeve, P.; Minor, P.D.; Schild, G.C.; Almond,
J.W.
#journal Eur. J. Biochem. (1983) 135:529-533
#title Nucleic acid sequence of the region of the genome encoding
capsid protein VP1 of neurovirulent and attenuated type 3
polioviruses
#cross-references MUID:84004370
#accession S42524
##molecule_type genomic RNA
##residues 579-878 #label STA
##cross-references EMBL:V01540; NID:g61153; PID:g929811
##experimental_source strain Sabin vaccine P3/Leon/37
COMMENT Coat proteins VP2 and VP3 and the RNA-directed RNA polymerase are
related to their counterparts in foot-and-mouth disease virus.
CLASSIFICATION #superfamily poliovirus genome polypeptide
KEYWORDS genome-linked protein; nucleotidyltransferase; phosphoprotein
FEATURE
2-69 #product coat protein VP4 #status predicted #label VP4\
70-340 #product coat protein VP2 #status predicted #label VP2\
341-578 #product coat protein VP3 #status predicted #label VP3\
579-878 #product coat protein VP1 #status predicted #label VP1\
879-1453 #product core protein P2-3b #status predicted #label
P23\
1027-1453 #product core protein P2-5b #status predicted #label
P25\
1124-1453 #product core protein P2-X #status predicted #label P2X\
1454-2206 #product protein P3-1b #status predicted #label P31\
1541-1562 #product genome-linked protein VPg #status predicted
#label VPg\
1563-2206 #product protein P3-2 #status predicted #label P32\
1563-1745 #product probable proteinase P3-7c #status predicted
#label P37\
1746-2206 #product RNA-directed RNA polymerase P3-4b #status
predicted #label P34\
1543 #binding_site phosphoryl-RNA (Tyr) (covalent) #status
predicted
SUMMARY #length 2206 #molecular-weight 246163 #checksum 7111
Query Match 76.0%; Score 79; DB 1; Length 2206;
Best Local Similarity 69.2%; Pred. No. 2.49e-03;
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
Db 679 KLFAMWKIIVKDT 691
QY 1 KFLAVWKIIVKDT 13
RESULT 8
ENTRY #type complete
TITLE genome polypeptide - human poliovirus 3 (strain Leon 12a1b)
CONTAINS coat protein VP1; coat protein VP2; coat protein VP3; coat
protein VP4; core protein P2-3b; core protein P2-5b; core
protein P2-X; genome-linked protein VPg; protein P3-1b;
protein P3-2; protein P3-4a; protein P3-6a; protein P3-6b;
protein P3-9; proteinase P3-7c; RNA-directed RNA polymerase
(EC 2.7.7.48) P3-4b
ORGANISM #formal_name human poliovirus 3
DATE 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change
ACCESSION S03822; S04920
REFERENCE S03822
#authors Toyoda, H.; Kohara, M.; Kataoka, Y.; Suganuma, T.; Omata, T.;
Imura, N.; Nomoto, A.
#journal J. Mol. Biol. (1984) 174:561-585
#title Complete nucleotide sequences of all three poliovirus
serotype genomes. Implication for genetic relationship,
gene function and antigenic determinants.
#cross-references MUID:84216300
#accession S03822
##molecule_type genomic RNA
##residues 1-2206 #label TOY
##cross-references EMBL:X00596
REFERENCE S04920
#authors Nomoto, A.
#submission submitted to the EMBL Data Library, April 1985
#accession S04920
##molecule_type genomic RNA
##residues 1-541, O', 543-696, M', 698-1311, E', 1313-2206 #label NOM
##cross-references EMBL:X00596; NID:g61139; PID:g61140
CLASSIFICATION #superfamily poliovirus genome polypeptide
KEYWORDS genome-linked protein; nucleotidyltransferase;

```

```

FEATURE
1543      #binding_site phosphoryl-RNA (Tyr) (covalent) #status
          predicted
SUMMARY   #length 2206 #molecular-weight 246276 #checksum 5631
          75.0%; Score 79; DB 2; Length 2206;
          Best Local Similarity 69.2%; Pred. No. 2.49e-03;
          Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 679 KLFAMWRITTKDT 691
QY 1 KFLAVWKITYKDT 13

RESULT 9
ENTRY   GNNY27
TITLE   genome polyprotein - human poliovirus 3 (strain 23127)
CONTAINS
protein VP1; coat protein VP2; coat protein VP3; coat
protein VP4; core protein P2-3b; core protein P2-5b; core
protein P2-X; genome-linked protein VPg; protein P3-1b;
proteinase; RNA-directed RNA polymerase 3
#formal_name human poliovirus 3
#formal_name human poliovirus 3
#length 2206 #sequence_revision 31-Dec-1988 #text_change
05-Jun-1998
ACCESSIONS
REFERENCE A27245
#authors Hughes, P.J.; Evans, D.M.A.; Minor, P.D.; Schild, G.C.;
          Almond, J.W.; Stanway, G.
#journal J. Gen. Virol. (1986) 67:2093-2102
#title The nucleotide sequence of a type 3 poliovirus isolated
          during a recent outbreak of poliomyelitis in Finland.
#cross-references MUID:87010550
#accession A27245
#molecule_type genomic RNA
#residues 1-2206 #label HUG
#cross-references GB:X04468; NID:g61112; PID:g61113
CLASSIFICATION #superfamily poliovirus genome polyprotein
KEYWORDS coat protein; core protein; genome-linked protein;
          nucleotidyltransferase; phosphoprotein; polyprotein;
          proteinase
FEATURE
1-69      #product coat protein VP4 #status predicted #label VP4\
70-340    #product coat protein VP2 #status predicted #label VP2\
341-578   #product coat protein VP3 #status predicted #label VP3\
579-878   #product coat protein VP1 #status predicted #label VP1\
879-1027  #product core protein P2-3b #status predicted #label
          P3b\
1028-1124 #product core protein P2-5b #status predicted #label
          P5b\
1125-1453 #product core protein P2-X #status predicted #label P2X\
1454-1540 #product protein P3-1b #status predicted #label P1b\
1541-1562 #product genome-linked protein VPg #status predicted
          #label VPg\
1563-1745 #product proteinase #status predicted #label PTS\
1746-2206 #product RNA-directed RNA polymerase #status predicted
          #label RNS\
1543      #binding_site phosphoryl-RNA (Tyr) (covalent) #status
          predicted
SUMMARY   #length 2206 #molecular-weight 245731 #checksum 265
          74.0%; Score 77; DB 1; Length 2206;
          Best Local Similarity 69.2%; Pred. No. 5.92e-03;
          Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 679 KLFAMWRITTKDT 691
QY 1 KFLAVWKITYKDT 13

RESULT 10
ENTRY   A57638
TITLE   receptor tyrosine kinase egl-15 precursor - Caenorhabditis
          elegans
FEATURE
1543      #formal_name Caenorhabditis elegans
          #formal_name Caenorhabditis elegans
          08-Feb-1996 #sequence_revision 08-Feb-1996 #text_change
          17-Mar-1999
ACCESSIONS
REFERENCE A57638
#authors DeVore, D.L.; Horvitz, H.R.; Stern, M.J.
#journal Cell (1995) 83:611-620
#title An EGF receptor signaling pathway is required for the normal
          cell migrations of the sex myoblasts in Caenorhabditis
          elegans hermaphrodites.
#cross-references MUID:9609862
#accession A57638
#status preliminary
#molecule_type mRNA
#residues 1-1040 #label DEV
#cross-references GB:U039761; NID:g1079711; PID:g1079712
CLASSIFICATION #superfamily unassigned Ser/Thr or Tyr-specific protein
          kinases; protein kinase homology
KEYWORDS ATP
FEATURE
638-935   #domain protein kinase homology #label KIN\
646-654   #region protein kinase ATP-binding motif
SUMMARY   #length 1040 #molecular-weight 118955 #checksum 4313
          67.3%; Score 70; DB 2; Length 1040;
          Best Local Similarity 56.7%; Pred. No. 1.13e-01;
          Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 651 FGWVKATYKKT 662
QY 2 FLAVWKITYKDT 13

RESULT 11
ENTRY   D69182
TITLE   conserved hypothetical protein MTH62 - Methanobacterium
          thermoautotrophicum (strain Delta H)
#formal_name Methanobacterium thermoautotrophicum
05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
04-Sep-1998
ACCESSIONS
REFERENCE D69182
#authors Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.;
          Dubois, J.; Aldredge, T.; Bashirzadeh, R.; Blakely, D.;
          Cook, R.; Gilbert, K.; Harrison, D.; Hoang, L.; Keagle, P.;
          Lumm, W.; Pothier, B.; Qiu, D.; Spadafora, R.; Vicaire, R.;
          Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiواني, N.; Caruso,
          A.; Bush, D.; Safer, H.; Patwell, D.; Prabhakar, S.;
          McDougall, S.; Shimer, G.; Goyal, A.; Pietrovskii, S.;
          Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling,
          J.; Reeve, J.N.
#journal J. Bacteriol. (1997) 179:7135-7155
#title Complete genome sequence of Methanobacterium
          thermoautotrophicum Delta H: functional analysis and
          comparative genomics.
#cross-references MUID:98037514
#accession D69182
#status preliminary; nucleic acid sequence not shown;
          translation not shown
#molecule_type DNA
#residues 1-250 #label MTH
#cross-references GB:AE000798; GB:AE000666; NID:g2621094; PID:g2621099
#experimental_source strain Delta H
GENETICS
#gene MTH62
#start_codon GTG
CLASSIFICATION #superfamily hypothetical protein M71632
SUMMARY   #length 250 #molecular-weight 28677 #checksum 3266
          64.4%; Score 67; DB 2; Length 250;
          Best Local Similarity 50.0%; Pred. No. 3.82e-01;
          Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

```

```

Db 187 FLSWVDVPIRIT 198
||:| | | | |
QY 2 FLAVWKITYKDT 13

RESULT 12
ENTRY #type complete
TITLE genome polyprotein - coxsackievirus A21 (strain Coe)
CONTAINS coat protein 1A; coat protein 1B; coat protein 1C; coat
protein 1D; core protein 2A; core protein 2B; core protein
2C; genome-linked protein VPg; protein 3A; proteinase;
RNA-directed RNA polymerase (EC 2.7.7.48)
ORGANISM #formal_name coxsackievirus A21
DATE 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change
05-Jun-1998
A33373
A33373
REFERENCE Hughes, P.J.; North, C.; Minor, P.D.; Stanway, G.
#authors J. Gen. Virol. (1989) 70:2943-2952
#journal The complete nucleotide sequence of coxsackievirus A21.
#title #cross-references MUID:90063544
#accession A33373
#molecule_type genomic RNA
#residues 1-2206 #label HUG
#cross-references GB:D00538; NID:g221147; PID:dl000880; PID:g221148
CLASSIFICATION #superfamily poliovirus genome polyprotein
KEYWORDS coat protein; core protein; genome-linked protein;
nucleotidyltransferase; phosphoprotein; polyprotein;
proteinase
FEATURE
1-69 #product coat protein 1A #status predicted #label VP1\
70-341 #product coat protein 1B #status predicted #label VP2\
342-578 #product coat protein 1C #status predicted #label VP3\
579-881 #product coat protein 1D #status predicted #label VP1\
882-1028 #product core protein 2A #status predicted #label PA2\
1029-1125 #product core protein 2B #status predicted #label PB2\
1126-1453 #product core protein 2C #status predicted #label PC2\
1454-1540 #product protein 3A #status predicted #label PA3\
1541-1562 #product genome-linked protein VPg #status predicted
#label PB3\
1563-1745 #product proteinase #status predicted #label PC3\
1746-2206 #product RNA-directed RNA polymerase #status predicted
#label PD3\
1543 #binding_site phosphoryl-RNA (Tyr) (covalent) #status
predicted
SUMMARY #length 2206 #molecular-weight 246049 #checksum 3669
Query Match 61.5%; Score 64; DB 1; Length 2206;
Best Local Similarity 53.8%; Pred. No. 1.25e+00;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 679 KHFNWNIYIDT 591
| : | : | | | |
QY 1 KFLAVWKITYKDT 13

RESULT 13
ENTRY #type complete
TITLE DNA polymerase III alpha chain (dnaE) homolog - Mycoplasma
genitalium (SGC3)
ORGANISM #formal_name Mycoplasma genitalium
DATE 17-Nov-1995 #sequence_revision 17-Nov-1995 #text_change
21-Aug-1998
H64228
A64260
A64260
REFERENCE Fraser, C.M.; Cocayne, J.D.; White, O.; Adams, M.D.; Clayton,
R.A.; Fleischmann, R.D.; Bult, C.J.; Kerlavage, A.R.;
Sutton, G.; Kelley, J.M.; Fritchman, J.L.; Weidman, J.F.;
Small, K.V.; Sandusky, M.; Fuhrmann, J.; Nguyen, D.;
Utterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick,
J.M.; Tomb, J.F.; Dougherty, B.A.; Bott, K.F.; Hu, P.C.;
Lucier, T.S.; Peterson, S.N.; Smith, H.O.; Hutchison III,
C.A.; Venter, J.C.
#journal Science (1995) 270:397-403
#title The minimal gene complement of Mycoplasma genitalium.
#cross-references MUID:96026346
#accession H64228
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-874 #label TIGR
#cross-references GB:U39705; GB:L43967; NID:g1045953; PID:g1045954;
TIGR:MG261
#experimental_source strain G-37
GENETICS
#genetic_code SGC3
CLASSIFICATION #superfamily polydeoxyribonucleotide synthase (NAD+)
SUMMARY #length 874 #molecular-weight 100434 #checksum 6779
Query Match 59.6%; Score 62; DB 2; Length 874;
Best Local Similarity 70.0%; Pred. No. 2.73e+00;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 118 LTTWKSTYKD 127
| : | | | | |
QY 3 LAVWKITYKD 12

RESULT 14
ENTRY #type complete
TITLE hemoglobin - southern lamprey
ORGANISM #formal_name Mordacia mordax #common_name southern lamprey
DATE 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change
10-Jul-1998
S13460
S13460
REFERENCE Hombados, I.; Vidal, Y.; Rodewald, K.; Braunitz, G.;
Neuzil, E.
#authors Biol. Chem. Hoppe-Seyler (1991) 372:49-56
#journal The primary structure of the hemoglobins of a southern
hemisphere lamprey (Mordacia mordax, Cyclostomata).
#title #cross-references MUID:91248417
#accession S13460
#molecule_type protein
#residues 1-149 #label HOM
FUNCTION #description in erythrocytes binds and transports molecular oxygen from
lung to tissues
CLASSIFICATION #superfamily globin; globin homolog
KEYWORDS chromoprotein; erythrocyte; heme; iron; oxygen carrier
FEATURE
11-149 #domain globin homology #label GLB\
73 #binding_site oxygen (His) (distal axial ligand) #status
predicted\
105 #binding_site heme iron (His) (proximal axial ligand)
#status predicted
SUMMARY #length 149 #molecular-weight 16508 #checksum 7504
Query Match 58.7%; Score 61; DB 2; Length 149;
Best Local Similarity 50.0%; Pred. No. 4.00e+00;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 18 KILAAMDIVYKN 29
| : | | | | |
QY 1 KFLAVWKITYKD 12

RESULT 15
ENTRY #type complete
TITLE hypothetical protein MJ1632 - Methanococcus jannaschii
ORGANISM #formal_name Methanococcus jannaschii
DATE 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change
04-Sep-1998
F64503
A64300
A64300
REFERENCE

```

```

#authors      Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann,
               R.D.; Sutton, G.G.; Blake, J.A.; Fitzgerald, L.M.; Clayton,
               R.A.; Gocayne, J.D.; Kerlavage, A.R.; Dougherty, B.A.;
               Tomb, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.;
               Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
               Scott, J.L.; Geoghegan, N.S.M.; Weidman, J.F.; Fuhrmann,
               J.L.; Nguyen, D.; Utterback, T.R.; Kelley, J.M.; Peterson,
               J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts,
               K.M.; Hurst, M.A.; Kaine, B.P.; Borodovsky, M.; Klenk,
               H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.
#journal      Science (1996) 273:1058-1073
#title        Complete genome sequence of the methanogenic archaeon,
               Methanococcus jannaschii.
#cross-references MVID:96337999
#accession    F64503
#status       preliminary; nucleic acid sequence not shown;
               translation not shown
#molecule_type DNA
#residues     1-255 #label BUL
#cross-references GB:U67603; GB:L77117; NID:g1592220; PID:g1592222;
               TIGR:MJ1632; PID:g1511593
GENETICS
#map_position FOR1612683-1613450
#start_codon TTG
CLASSIFICATION #superfamily hypothetical protein MJ1632
SUMMARY        #length 255 #molecular-weight 29795 #checksum 9201

Query Match      58.7%; Score 61; DB 2; Length 255;
Best Local Similarity 46.2%; Pred. No. 4.00e+00;
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Db 193 KFLSNWDLPYRIT 205
   |||:|:|:|
QY 1 KFLAVWKIIYKDT 13

```

Search completed: Tue Aug 17 16:03:06 1999
 Job time : 20 secs.

This Page Blank (uspto)

WORLD

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Tue Aug 17 16:03:24 1999; MasPar time 2.99 Seconds
Tabular output not generated. 122.872 Million cell updates/sec

Title: >US-09-049-847-4
Description: (1-13) from US09049847.pep
Perfect Score: 104
Sequence: 1 KFLAVWKITYKDT 13

Scoring table: PAM 150
Gap 15

Searched: 77977 seqs, 28266293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 27.515; Variance 38.609; scale 0.713

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	87	83.7	2206	1	POLG_POL1M GENOME POLYPROTEIN [CO	7.73e-06
2	87	83.7	2208	1	POLH_POL1M GENOME POLYPROTEIN [CO	7.73e-06
3	85	81.7	2205	1	POLG_POL2W GENOME POLYPROTEIN [CO	2.11e-05
4	85	81.7	2207	1	POLG_POL2L GENOME POLYPROTEIN [CO	2.11e-05
5	85	81.7	2209	1	POLG_POL1S GENOME POLYPROTEIN [CO	2.11e-05
6	79	76.0	2206	1	POLG_POL1L GENOME POLYPROTEIN [CO	4.01e-04
7	77	74.0	2206	1	POLG_POL32 GENOME POLYPROTEIN [CO	1.05e-03
8	70	67.3	1040	1	EG15_CAEEL MYOBLAST GROWTH FACTOR	2.74e-02
9	64	61.5	2206	1	POLG_CXA21 GENOME POLYPROTEIN [CO	3.91e-01
10	62	59.6	874	1	DP3A_MYCGE DNA POLYMERASE III, AL	9.20e-01
11	61	58.7	149	1	GLB3_MORMR GLOBIN III.	1.40e+00
12	59	56.7	1938	1	MYSS_CHICK MYOSIN HEAVY CHAIN, SK	3.21e+00
13	58	55.8	230	1	YX09_CAEEL HYPOPHETICAL 26.5 KD P	4.81e+00
14	58	55.8	847	1	SYA_HELPY ALANYL-TRNA SYNTHETASE	4.81e+00
15	58	55.8	874	1	POLL_HUMAN RETROVIRUS-RELATED POL	4.81e+00
16	58	55.8	878	1	SYA_BACSU ALANYL-TRNA SYNTHETASE	4.81e+00
17	58	55.8	1332	1	XKDO_BACSU PHAGE-LIKE ELEMENT PBS	4.81e+00
18	57	54.8	241	1	YBN0_YEAST HYPOPHETICAL 27.5 KD P	7.19e+00
19	57	54.8	396	1	RMAR_YEAST MITOCHONDRIAL RIBOSOMA	7.19e+00
20	57	54.8	415	1	EF1G_YEAST ELONGATION FACTOR 1-GA	7.19e+00
21	57	54.8	467	1	KCOT_MOUSE COT PROTO-ONCOGENE SER	7.19e+00
22	57	54.8	644	1	VP4_BT11 VP4 CORE PROTEIN.	7.19e+00
23	57	54.8	644	1	VP4_BT13 VP4 CORE PROTEIN.	7.19e+00

24	57	54.8	654	1	VP4_BT10 VP4 CORE PROTEIN.	7.19e+00
25	57	54.8	1940	1	MYSE_CHICK MYOSIN HEAVY CHAIN, FA	7.19e+00
26	56	53.8	151	1	CAV3_HUMAN CAVEOLIN-3 (M-CAVEOLIN	1.07e-01
27	56	53.8	190	1	YCG8_YEAST VERY HYPOPHETICAL 22.1	1.07e-01
28	56	53.8	346	1	YRAQ_ECOLI HYPOPHETICAL 37.3 KD P	1.07e-01
29	56	53.8	467	1	KCOT_HUMAN COT PROTO-ONCOGENE SER	1.07e-01
30	56	53.8	538	1	BUD5_YEAST BUD 5LIE SELECTION PRO	1.07e-01
31	56	53.8	904	1	SYA_MYCTU ALANYL-TRNA SYNTHETASE	1.07e-01
32	56	53.8	1934	1	MISB_MESAU MYOSIN HEAVY CHAIN, CA	1.07e-01
33	56	53.8	1935	1	MYSB_HUMAN MYOSIN HEAVY CHAIN, CA	1.07e-01
34	56	53.8	1935	1	MYSB_PIG MYOSIN HEAVY CHAIN, CA	1.07e-01
35	56	53.8	1938	1	MISA_MOUSE MYOSIN HEAVY CHAIN, CA	1.07e-01
36	56	53.8	1938	1	MISA_RAT MYOSIN HEAVY CHAIN, CA	1.07e-01
37	56	53.8	1939	1	MISA_MESAU MYOSIN HEAVY CHAIN, CA	1.07e-01
38	56	53.8	1939	1	MISA_HUMAN MYOSIN HEAVY CHAIN, CA	1.07e-01
39	56	53.8	2329	1	YS89_CAEEL HYPOPHETICAL 254.3 KD	1.07e-01
40	56	53.8	4196	1	DYHC_SCHPO DYNEIN HEAVY CHAIN, CY	1.07e-01
41	55	52.9	46	1	YPC4_ECOLI HYPOPHETICAL 5.3 KD PR	1.58e-01
42	55	52.9	240	1	RESD_BACSD TRANSCRIPTIONAL REGULA	1.58e-01
43	55	52.9	715	1	PERE_HUMAN EOSINOPHIL PEROXIDASE	1.58e-01
44	55	52.9	860	1	SYL_ECOLI LEUCYL-TRNA SYNTHETASE	1.58e-01
45	55	52.9	1096	1	PULA_KUEAE PULLULANASE PRECURSOR	1.58e-01

ALIGNMENTS

RESULT	1		
ID	POLG_POLIM	STANDARD;	PRT; 2206 AA.
AC	P03299;		
DT	21-JUL-1986 (REL. 01, CREATED)		
DT	01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)		
DT	15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)		
DE	GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS		
DE	P2A TO P2C, P3A; GENOME-LINKED PROTEIN VPg; PICORNAIN 3C		
DE	(EC 3.4.22.28) (PROTEASE 3C) (P3C); RNA-DIRECTED RNA POLYMERASE P3D		
DE	(EC 2.7.7.48)] (VERSION 1).		
OS	POLIOVIRUS TYPE 1 (STRAIN MAHONEY).		
OC	VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;		
OC	ENTEROVIRUS.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE; 81220953.		
RA	KITAMURA N., SEMLER B.L., ROTHBERG P.G., LARSEN G.R., ADLER C.J.,		
RA	DORNER A.J., EMINI E.A., HANECAK R., LEE J.J., VAN DER WERF S.,		
RA	ANDERSON C.W., WIMMER E.;		
RT	"Primary structure, gene organization and polypeptide expression of		
RT	poliovirus RNA.";		
RL	NATURE 291:547-553(1981).		
RN	[2]		
RP	SEQUENCE OF 1538-1573 FROM N.A.		
RX	MEDLINE; 81001866.		
RA	KITAMURA N., ADLER C.J., ROTHBERG P.G., MARTINKO J., NATHENSON S.G.,		
RA	WIMMER E.;		
RT	"The genome-linked protein of picornaviruses. VII. Genetic mapping of		
RT	poliovirus VPg by protein and RNA sequence studies.";		
RL	CELL 21:295-302(1980).		
RN	[3]		
RP	SEQUENCE OF 1-68.		
RX	MEDLINE; 82242310.		
RA	DORNER A.J., DORNER L.F., LARSEN G.R., WIMMER E., ANDERSON C.W.;		
RT	"Identification of the initiation site of poliovirus polyprotein		
RT	synthesis.";		
RL	J. VIROL. 42:1017-1028(1982).		
RN	[4]		
RP	X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS) OF 1-880.		
RX	MEDLINE; 85300512.		
RA	HOGLE J.M., CHOW M., FILMAN D.J.;		
RT	"Three-dimensional structure of poliovirus at 2.9-A resolution.";		
RL	SCIENCE 229:1358-1365(1985).		
RN	[5]		
RP	X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS) OF 1-880.		
RX	MEDLINE; 95120467.		
RA	GRANT R.A., HIREMATH C.N., FILMAN D.J., SYED R., ANDRIES K.		

```

RA HOGLE J.M.;
RT "Structures of poliovirus complexes with anti-viral drugs:
RL implications for viral stability and drug design.";
CC CURR. BIOL. 4:784-787(1994).
CC -!- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN
CC Q/G SITES IN THE POLYPROTEIN. IT MAY BE A CYSTEINE PROTEASE.
CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -!- PTM: VPG IS LINKED, PROBABLY BY TYR-1545, TO THE URIDYLATE RESIDUE
CC AT THE 5' END OF THE GENOME RNA. IT MAY PLAY A ROLE IN THE
CC INITIATION OF RNA SYNTHESIS AND IT MAY ALSO BE INVOLVED IN
CC MORPHOGENESIS.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -!- CAUTION: SEE ALSO VERSION 2 OF THIS PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; V01148; G61237; -
DR PIR; A03897; GNNYLP.
DR PDB; 2BLV; 15-JUL-93.
DR PDB; 1EFT; 31-MAR-95.
DR PDB; 1PCV; 07-DEC-95.
DR PDB; 1VBD; 11-JUL-96.
DR PDB; 1AL2; 19-NOV-97.
DR PDB; 1A86; 03-DEC-97.
DR PDB; 1A87; 03-DEC-97.
DR PDB; 1A88; 03-DEC-97.
DR PDB; 1A89; 03-DEC-97.
DR PDB; 1ASJ; 03-DEC-97.
DR PDB; 1P01; 03-DEC-97.
DR PDB; 1P02; 03-DEC-97.
DR PFAM; PF00073; rnv; 3.
DR PFAM; PF00548; Cys-protease-3C; 1.
DR PFAM; PF00680; RNA_dep_RNA_pol; 1.
DR PFAM; PF00910; RNA_helicase; 1.
DR PFAM; PF00947; Pico_P2A; 1.
KW POLYPROTEIN; COAT PROTEIN; CORE PROTEIN; TRANSFERASE;
KW RNA-DIRECTED RNA POLYMERASE; HYDROLASE; THIOL PROTEASE; MYRISTYLATION;
KW 3D-STRUCTURE.
FT INIT_MET 0 0
FT CHAIN 1 68 COAT PROTEIN VP4.
FT CHAIN 69 339 COAT PROTEIN VP2.
FT CHAIN 340 577 COAT PROTEIN VP3.
FT CHAIN 578 879 COAT PROTEIN VP1.
FT CHAIN 880 1028 CORE PROTEIN P2A.
FT CHAIN 1029 1125 CORE PROTEIN P2B.
FT CHAIN 1126 1454 CORE PROTEIN P2C.
FT CHAIN 1455 1541 CORE PROTEIN P3A.
FT CHAIN 1542 1563 GENOME-LINKED PROTEIN VPG.
FT CHAIN 1564 1745 PICORNAIN 3C.
FT CHAIN 1746 2206 RNA-DIRECTED RNA POLYMERASE P3D.
FT LIPID 1 1 MYRISTATE.
FT BINDING 1544 1544 URIDYLATE AT THE 5' END OF THE GENOME RNA.
FT ACT_SITE 1710 1710 PROTEASE (POTENTIAL).
FT STRAND 3 6
FT STRAND 25 28
FT STRAND 35 37
FT HELIX 45 45
FT STRAND 49 49
FT TURN 49 49
FT HELIX 50 53
FT STRAND 56 56
FT TURN 62 63
FT TURN 74 75
FT TURN 79 80
FT STRAND 82 86

```

```

FT TURN 87 88
FT STRAND 89 93
FT STRAND 96 101
FT HELIX 102 104
FT TURN 112 114
FT STRAND 122 122
FT HELIX 125 127
FT TURN 128 128
FT STRAND 132 133
FT STRAND 137 140
FT TURN 141 142
FT STRAND 146 150
FT TURN 151 151
FT HELIX 152 154
FT TURN 155 156
FT HELIX 158 166
FT STRAND 167 179
FT TURN 184 185
FT STRAND 186 196
FT TURN 197 197
FT STRAND 202 202
FT HELIX 213 216
FT HELIX 219 221
FT STRAND 223 224
FT STRAND 226 226
FT TURN 235 236
FT STRAND 242 242
FT STRAND 245 245
FT HELIX 246 248
FT TURN 249 252
FT STRAND 255 259
FT HELIX 262 266
FT TURN 267 269
FT STRAND 272 277
FT STRAND 286 286
FT TURN 288 290
FT STRAND 291 291
FT STRAND 294 306
FT STRAND 313 330
FT STRAND 336 336
FT TURN 347 350
FT TURN 354 355
FT STRAND 362 362
FT TURN 365 366
FT STRAND 378 379
FT STRAND 381 381
FT HELIX 382 386
FT TURN 387 387
FT STRAND 390 391
FT STRAND 396 396
FT TURN 398 402
FT HELIX 404 407
FT STRAND 409 412
FT STRAND 422 425
FT TURN 428 430
FT TURN 432 436
FT HELIX 438 443
FT TURN 444 445
FT STRAND 446 450
FT STRAND 453 459
FT TURN 463 464
FT STRAND 466 466
FT STRAND 468 474
FT HELIX 484 488
FT TURN 489 489
FT STRAND 491 496
FT STRAND 502 507
FT STRAND 516 517
FT HELIX 522 524
FT STRAND 528 533
FT TURN 538 538
FT TURN 541 542
FT STRAND 546 555

```

FT TURN 557 558
 FT STRAND 560 564
 FT STRAND 599 599
 FT TURN 600 601
 FT STRAND 603 603
 FT STRAND 607 607
 FT TURN 619 620
 FT STRAND 621 622

...
 Note: remainder of annotations omitted.

Query Match 83.7%; Score 87; DB 1; Length 2206;
 Best Local Similarity 84.6%; Pred.No. 7.73e-06;
 Matches 11; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 680 KLFVAVKITYKDT 692
 I:|||||
 QY 1 KLFVAVKITYKDT 13

RESULT 2
 ID POLH_POLIM STANDARD; PRT: 2208 AA.

AC P03300; Q84879; Q84880; Q89679;
 DT 21-JUL-1986 (REL. 01, CREAIED)
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
 DE GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS
 DE P2A TO P2C, P3A; GENOME-LINKED PROTEIN VPG; PICORNAIN 3C
 DE (EC 3.4.22.28) (PROTEASE 3C) (P3C); RNA-DIRECTED RNA POLYMERASE P3D
 DE (EC 2.7.7.48)] (VERSION 2)
 OS POLIOVIRUS TYPE 1 (STRAIN MAHONEY).
 OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;
 OC ENTEROVIRUS.

EN [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE: 82060159.
 RA RACANTIELLO V.R., BALTIMORE D.;
 RT "Molecular cloning of poliovirus cDNA and determination of the
 RT complete nucleotide sequence of the viral genome.";
 RL PROC. NATL. ACAD. SCI. U.S.A. 78:4887-4891(1981).
 RN [2]
 RX "Three-dimensional structure of poliovirus at 2.9-A resolution.";
 RL SCIENCE 229:1358-1365(1985).
 RN [3]

RX X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS) OF 1-880.
 RX MEDLINE: 95120467.
 RA GRANT R.A., HIREMATH C.N., FILMAN D.J., SYED R., ANDRIES K.,
 RA HOGLE J.M.;
 RT "structures of poliovirus complexes with anti-viral drugs:
 RT implications for viral stability and drug design.";
 RL CURR. BIOL. 4:784-797(1994).

CC Q/G SITES IN THE POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN
 CC FUNCTION: P3C POLYPEPTIDE. IT MAY BE A CYSTEINE PROTEASE.

CC SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
 CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
 CC VP3, AND VP4.
 CC -!- PTM: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD MATURE PROTEINS.

CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
 CC -!- CAUTION: SEE ALSO VERSION 1 OF THIS PROTEIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 CC EMBL: J02281; G332887; -;
 CC EMBL: V01149; G61253; -;

DR EMBL; V01149; E9211; ALT_SEQ.
 DR EMBL; V01149; E9212; ALT_SEQ.
 DR EMBL; V01149; E9213; ALT_SEQ.
 DR PIR; A03898; GNNY2P.
 DR PDB; 2PLV; 15-JUL-93.
 DR PDB; 1FPT; 31-MAR-95.
 DR PDB; 1POV; 07-DEC-95.
 DR PDB; 1VBD; 11-JUL-96.
 DR PDB; 1AL2; 19-NOV-97.
 DR PDB; 1AR6; 03-DEC-97.
 DR PDB; 1AR7; 03-DEC-97.
 DR PDB; 1AR8; 03-DEC-97.
 DR PDB; 1AR9; 03-DEC-97.
 DR PDB; 1ASJ; 03-DEC-97.
 DR PDB; 1PO1; 03-DEC-97.
 DR PDB; 1PO2; 03-DEC-97.
 DR PFAM; PF00073; rhv; 3.
 DR PFAM; PF00548; Cys-protease-3C; 1.
 DR PFAM; PF00680; RNA_dep_RNA_pol; 1.
 DR PFAM; PF00910; RNA_helicase; 1.
 DR PFAM; PF00947; Pico_P2A; 1.
 KW POLYPROTEIN; COAT PROTEIN; CORE PROTEIN; TRANSFERASE;
 KW RNA-DIRECTED RNA POLYMERASE; HYDROLASE; THIOL PROTEASE; MYRISTYLATION;
 KW 3D-STRUCTURE.
 KW INIT_MET 0 0
 FT CHAIN 1 68 COAT PROTEIN VP4.
 FT CHAIN 69 340 COAT PROTEIN VP2.
 FT CHAIN 341 578 COAT PROTEIN VP3.
 FT CHAIN 579 880 COAT PROTEIN VP1.
 FT CHAIN 881 1029 CORE PROTEIN P2A.
 FT CHAIN 1030 1126 CORE PROTEIN P2B.
 FT CHAIN 1127 1453 CORE PROTEIN P2C.
 FT CHAIN 1456 1542 CORE PROTEIN P3A.
 FT CHAIN 1543 1564 GENOME-LINKED PROTEIN VPG.
 FT CHAIN 1565 1746 PICORNAIN 3C.
 FT CHAIN 1747 2208 RNA-DIRECTED RNA POLYMERASE P3D.
 FT LIPID 1 1 MYRISTATE.
 FT ACT_SITE 1711 1711 PROTEASE (POTENTIAL).
 FT ACT_SITE 1725 1725 PROTEASE (POTENTIAL).
 FT STRAND 3 6
 FT STRAND 25 28
 FT HELIX 35 37
 FT STRAND 45 45
 FT TURN 49 49
 FT HELIX 50 53
 FT STRAND 56 56
 FT TURN 62 63
 FT TURN 74 75
 FT TURN 79 80
 FT STRAND 82 86
 FT TURN 87 88
 FT STRAND 89 93
 FT STRAND 96 101
 FT TURN 102 104
 FT TURN 112 114
 FT STRAND 122 122
 FT HELIX 125 127
 FT TURN 128 128
 FT STRAND 132 133
 FT STRAND 137 140
 FT TURN 141 142
 FT STRAND 146 150
 FT TURN 151 151
 FT HELIX 152 154
 FT TURN 155 156
 FT HELIX 158 166
 FT STRAND 167 179
 FT TURN 184 185
 FT STRAND 186 196
 FT TURN 197 197
 FT STRAND 202 202
 FT HELIX 213 216
 FT HELIX 219 221

FT ACT_SITE 1724 1724 PROTEASE (POTENTIAL).
SQ SEQUENCE 2205 AA; 245701 MW; DA976BE8 CRC32;

Query Match 81.7%; Score 85; DB 1; Length 2205;
Best Local Similarity 76.9%; Pred. No. 2.11e-05;
Matches 10; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 681 KLFVWKIYKDT 693
QY 1 KFLAVWKIYKDT 13

RESULT 4
ID POLG_POL2L STANDARD; PRT: 2207 AA.
AC P06210;
DI 01-JAN-1988 (REL. 06, CREATED)
DI 01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)
DI 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE GENOME POLYPEPTIDE [CONTAINS: COAT PROTEIN VP1 TO VP4; CORE PROTEINS P2A TO P2C, P3A; GENOME-LINKED PROTEIN VPG; PICORNAIN 3C
DE (EC 3.4.22.28) (PROTEASE 3C) (P3C); RNA-DIRECTED RNA POLYMERASE P3D
DE (EC 2.7.7.48)]
OS POLIOVIRUS TYPE 2 (STRAIN LANSING).
OC VIRUSES: SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;
OC ENTEROVIRUS.

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 86115399.
RA LA MONICA N., MERIAM C., RACANIELLO V.R.;
RT "Mapping of sequences required for mouse neurovirulence of poliovirus type 2 Lansing";
RL J. VIROL. 57:515-525(1986);

CC -!- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN Q/G SITES IN THE POLYPROTEIN. IT MAY BE A CYSTEINE PROTEASE.
CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS. EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2, VP3, AND VP4.

CC -!- PTM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----

CC EMBL; M12197; G332851; -
CC PIR; A29507; GNNY5P. -
CC PFAM; PF00073; rhv; 3.
CC PFAM; PF00548; Cys-protease-3c; 1.
CC PFAM; PF00680; RNA_dep_RNA_pol; 1.
CC PFAM; PF00910; RNA_helicase; 1.
CC PFAM; PF00947; Pico_P2A; 1.
CC HSP; P03299; IPOV.
CC POLYPROTEIN; COAT PROTEIN; CORE PROTEIN; TRANSFERASE;
KW RNA-DIRECTED RNA POLYMERASE; HYDROLASE; THIOL PROTEASE; MYRISTYLATION.
FT CHAIN 2 69
FT CHAIN 70 340
FT CHAIN 341 578
FT CHAIN 579 879
FT CHAIN 880 1028
FT CHAIN 1029 1125
FT CHAIN 1126 1454
FT CHAIN 1455 1541
FT CHAIN 1542 1563
FT CHAIN 1564 1746
FT CHAIN 1747 2207
FT CHAIN 2 2
FT LIPID 1710 1710
FT ACT_SITE 1724 1724
FT SEQUENCE 2207 AA; 245829 MW; 36A83606 CRC32;

Query Match 81.7%; Score 85; DB 1; Length 2207;
Best Local Similarity 76.9%; Pred. No. 2.11e-05;
Matches 10; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 681 KLFVWKIYKDT 693
QY 1 KFLAVWKIYKDT 13

RESULT 5
ID POLG_POL1S STANDARD; PRT: 2209 AA.
AC P03301;
DI 21-JUL-1986 (REL. 01, CREATED)
DI 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DI 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE GENOME POLYPEPTIDE [CONTAINS: COAT PROTEIN VP1 TO VP4; CORE PROTEINS P2A TO P2C, P3A; GENOME-LINKED PROTEIN VPG; PICORNAIN 3C
DE (EC 3.4.22.28) (PROTEASE 3C) (P3C); RNA-DIRECTED RNA POLYMERASE P3D
DE (EC 2.7.7.48)]
OS POLIOVIRUS TYPE 1 (STRAIN SABIN).
OC VIRUSES: SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;
OC ENTEROVIRUS.

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 8329876.
RA NOMOTO A., OMATA I., TOYODA H., KUGE S., HORIE H., KARAOKA Y.,
RA GENBA Y., NAKANO Y., IMURA N.;
RT "Complete nucleotide sequence of the attenuated poliovirus Sabin 1 strain genome";
RL PROC. NATL. ACAD. SCI. U.S.A. 79:5793-5797(1982);

CC -!- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN Q/G SITES IN THE POLYPROTEIN. IT MAY BE A CYSTEINE PROTEASE.
CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS. EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2, VP3, AND VP4.

CC -!- PTM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -!- THIS VIRUS IS A LIVE VACCINE STRAIN DERIVED FROM THE MAHONEY STRAIN BY SPONTANEOUS MUTATIONS DURING THE ATTENUATION PROCESS.

CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----

CC EMBL; V01150; G61258; -
CC PIR; A03899; GNNY3P. -
CC PFAM; PF00073; rhv; 3.
CC PFAM; PF00548; Cys-protease-3c; 1.
CC PFAM; PF00680; RNA_dep_RNA_pol; 1.
CC PFAM; PF00910; RNA_helicase; 1.
CC PFAM; PF00947; Pico_P2A; 1.
CC HSP; P03299; IPOV.
CC POLYPROTEIN; COAT PROTEIN; CORE PROTEIN; TRANSFERASE;
KW RNA-DIRECTED RNA POLYMERASE; HYDROLASE; THIOL PROTEASE; MYRISTYLATION.
FT CHAIN 2 69
FT CHAIN 70 341
FT CHAIN 342 579
FT CHAIN 580 861
FT CHAIN 862 1030
FT CHAIN 1031 1127
FT CHAIN 1128 1456
FT CHAIN 1457 1543
FT CHAIN 1544 1565
FT CHAIN 1566 1747
FT CHAIN 1748 2209
FT CHAIN 2 2
FT LIPID 1712 1712
FT ACT_SITE 1726 1726
FT SEQUENCE 2209 AA; 245829 MW; 36A83606 CRC32;

```
SQ SEQUENCE 2209 AA; 246576 MW; 25407F3A CRC32;
Query Match 81.78; Score 85; DB 1; Length 2209;
Best Local Similarity 76.98; Pred.No. 2.11e-05;
Matches 10; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 682 KLFVWKITYKDT 694
   1 KFLAVWKITYKDT 13

QY 1 KFLAVWKITYKDT 13

RESULT 6
ID POLG_POL3L STANDARD; PRT; 2206 AA.
AC P03302;
DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DE GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS
DE P2A TO P2C, P3A; GENOME-LINKED PROTEIN VP3; PICORNAIN 3C
DE (EC 3.4.22.28) (PROTEASE 3C) (P3C); RNA-DIRECTED RNA POLYMERASE P3D
DE (EC 2.7.7.48)].
OS POLIOVIRUS TYPE 3 (STRAINS P3/LEON/37 AND P3/LEON 12A[1]B).
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;
OC ENTEROVIRUS.
RN
RP SEQUENCE FROM N.A.
RC STRAIN=P3/LEON/37;
RX MEDLINE; 84170338;
RA STANWAY G., HUGHES P.J., MOUNTFORD R.C., REEVE P., MINOR P.D.,
RA SCHILD G.C., ALMOND J.W.;
RT "Comparison of the complete nucleotide sequences of the genomes of
RT the neurovirulent poliovirus P3/Leon/37 and its attenuated Sabin
RT vaccine derivative P3/Leon 12a1b."
RL PROC. NACL. ACAD. SCI. U.S.A. 81:1539-1543 (1984).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=P3/LEON 12A[1]B;
RX MEDLINE; 83299239;
RA STANWAY G., CANN A.J., HAUTPMANN R., HUGHES P., CLARKE L.D.,
RA MOUNTFORD R.C., MINOR P.D., SCHILD G.C., ALMOND J.W.;
RT "The nucleotide sequence of poliovirus type 3 leon 12 alb: comparison
RT with poliovirus type 1."
RL NUCLEIC ACIDS RES. 11:5629-5643 (1983).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS) OF 1-878.
RX MEDLINE; 95120467;
RA GRANT R.A., HIREMATH C.N., FILMAN D.J., SYED R., ANDRIES K.,
RA HOGUE J.M.;
RT "Structures of poliovirus complexes with anti-viral drugs:
RT implications for viral stability and drug design."
RL CURR. BIOL. 4:784-797 (1994).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1-878.
RA SYED R., FILMAN D.J., HOGUE J.M.;
RL SUBMITTED (MAR-1995) TO THE PDB DATA BANK.
CC -!- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN
CC O/G SITES IN THE POLYPROTEIN. IT MAY BE A CYSTEINE PROTEASE.
CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC -!- PTM: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -!- THE SEQUENCE OF STRAIN SABIN VACCINE P3/LEON/37 IS SHOWN.
CC -!- THE STRAIN SABIN VACCINE P3/LEON/37 IS THE PROGENITOR OF THE
CC STRAIN SABIN VACCINE P3/LEON 12A[1]B.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
CC EMBL; X00925; G61155; -
DR PIR; A03900; GNNV4p.
DR PDB; 1PIV; 20-JUL-95.
DR PDB; 1PVC; 15-SEP-95.
DR PDB; 1VEA; 11-JUL-96.
DR PDB; 1VBB; 11-JUL-96.
DR PDB; 1VBC; 11-JUL-96.
DR PDB; 1VBE; 11-JUL-96.
DR PFAM; PF00073; rnv; 3.
DR PFAM; PF00548; Cys-protease-3C; 1.
DR PFAM; PF00680; RNA_dep_RNA_pol; 1.
DR PFAM; PF00910; RNA_helicase; 1.
DR PFAM; PF00947; Pico_P2A; 1.
KW POLYPROTEIN; COAT PROTEIN; CORE PROTEIN; TRANSFERASE;
KW RNA-DIRECTED RNA POLYMERASE; HYDROLASE; THIOL PROTEASE; MYRISTYLATION;
KW 3D-STRUCTURE.
FT CHAIN 2 69 COAT PROTEIN VP4.
FT CHAIN 70 340 COAT PROTEIN VP2.
FT CHAIN 341 578 COAT PROTEIN VP3.
FT CHAIN 579 878 COAT PROTEIN VP1.
FT CHAIN 879 1027 CORE PROTEIN P2A.
FT CHAIN 1028 1124 CORE PROTEIN P2B.
FT CHAIN 1125 1453 CORE PROTEIN P2C.
FT CHAIN 1454 1540 CORE PROTEIN P3A.
FT CHAIN 1541 1562 GENOME-LINKED PROTEIN VP3.
FT CHAIN 1563 1745 PICORNAIN 3C.
FT CHAIN 1746 2206 MYRISTATE.
FT LIPID 2 2 MYRISTATE.
FT ACT_SITE 1709 1709 PROTEASE (POTENTIAL).
FT ACT_SITE 1723 1723 PROTEASE (POTENTIAL).
FT VARIANT 431 431 S -> F (IN P3/LEON 12A[1]B).
FT VARIANT 864 864 K -> R (IN P3/LEON 12A[1]B).
FT VARIANT 908 908 T -> A (IN P3/LEON 12A[1]B).
SQ SEQUENCE 2206 AA; 246163 MW; 78B79E4F CRC32;

Query Match 76.08; Score 79; DB 1; Length 2206;
Best Local Similarity 69.28; Pred.No. 4.01e-04;
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 679 KLFAMWKITYKDT 691
   1 KFLAVWKITYKDT 13

QY 1 KFLAVWKITYKDT 13

RESULT 7
ID POLG_POL32 STANDARD; PRT; 2206 AA.
AC P06209;
DT 01-JAN-1988 (REL. 06, CREATED)
DT 01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS
DE P2A TO P2C, P3A; GENOME-LINKED PROTEIN VP3; PICORNAIN 3C
DE (EC 3.4.22.28) (PROTEASE 3C) (P3C); RNA-DIRECTED RNA POLYMERASE P3D
DE (EC 2.7.7.48)].
OS POLIOVIRUS TYPE 3 (STRAIN 23127).
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;
OC ENTEROVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 87010550.
RA HUGHES P.J., EVANS D.M.A., MINOR P.D., SCHILD G.C., ALMOND J.W.,
RA STANWAY G.;
RT "The nucleotide sequence of a type 3 poliovirus isolated during a
RT recent outbreak of poliomyelitis in Finland."
RL J. GEN. VIROL. 67:2093-2102 (1986).
CC -!- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN
CC O/G SITES IN THE POLYPROTEIN. IT MAY BE A CYSTEINE PROTEASE.
CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC -!- PTM: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
```

```

CC CC      -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC CC      -----
CC CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC      the European Bioinformatics Institute. There are no restrictions on its
CC CC      use by non-profit institutions as long as its content is in no way
CC CC      modified and this statement is not removed. Usage by and for commercial
CC CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC      or send an email to license@isb-sib.ch).
CC CC      -----
CC CC      EMBL; X04468; G61113; -.
CC CC      PIR; A27245; GNIN27.
CC CC      PFAM; PF00073; rhv; 3.
CC CC      PFAM; PF00548; Cys-protease-3C; 1.
CC CC      PFAM; PF00680; RNA_dep_RNA_pol; 1.
CC CC      PFAM; PF00910; RNA_helicase; 1.
CC CC      PFAM; PF00947; Pico_P2A; 1.
CC CC      HSP; P03302; IPIV.
CC CC      KW POLYPROTEIN; COAT PROTEIN; CORE PROTEIN; TRANSFERASE;
CC CC      RNA-DIRECTED RNA POLYMERASE; HYDROLASE; THIOL PROTEASE; MYRISTYLATION.
CC CC      FT CHAIN 2 69
CC CC      FT CHAIN 70 340
CC CC      FT CHAIN 341 578
CC CC      FT CHAIN 579 878
CC CC      FT CHAIN 879 1027
CC CC      FT CHAIN 1028 1124
CC CC      FT CHAIN 1125 1453
CC CC      FT CHAIN 1454 1540
CC CC      FT CHAIN 1541 1562
CC CC      FT CHAIN 1563 1745
CC CC      FT CHAIN 1746 2206
CC CC      FT LIPID 2 2
CC CC      FT ACT_SITE 1709 1709
CC CC      FT ACT_SITE 1723 1723
CC CC      SQ SEQUENCE 2206 AA; 245731 MW; 245731 MW; DD5809F5 CRC32;

Query Match      74.0%; Score 77; DB 1; Length 2206;
Best Local Similarity 69.2%; Pred.No. 1.05e-03;
Matches      9; Conservative      3; Mismatches      1; Indels      0; Gaps      0;

Db      679 KLFATWRTYKDT 691
QY      1 KFLAVWKITYKDT 13

RESULT      8
ID      EG15-CABEL      STANDARD;      PRT; 1040 AA.
AC      Q10656;
DT      01-NOV-1997 (REL. 35, CREATED)
DT      01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DE      MYOBLAST GROWTH FACTOR RECEPTOR ELG-15 PRECURSOR (EC 2.7.1.112).
GN      EGL-15.
OS      CAENORHABDITIS ELEGANS.
OC      EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC      RHABDITINA; RHABDITODEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-BRISTOL N2.
RX      MEDLINE; 96069862.
RA      DEVORE D.L., HORVITZ H.R., STERN M.J.;
RT      "An EGF receptor signaling pathway is required for the normal cell
RL      migrations of the sex myoblasts in C. elegans hermaphrodites.";
CC      CELL 83:611-620(1995).
CC      -!- FUNCTION: RECEPTOR FOR AN EXTRACELLULAR SIGNAL REQUIRED FOR
CC      THE GUIDANCE OF SEX MYOBLAST MIGRATION DURING GONAD DEVELOPMENT.
CC      -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC      -!- SIMILARITY: BELONGS TO THE FIBROBLAST GROWTH FACTOR RECEPTOR
CC      FAMILY.
CC      -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. THE
CC      EXTRACELLULAR DOMAIN CONTAINS 3 IG-LIKE DOMAINS.
CC CC      -----
CC CC      This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC      the European Bioinformatics Institute. There are no restrictions on its
CC CC      use by non-profit institutions as long as its content is in no way
CC CC      modified and this statement is not removed. Usage by and for commercial
CC CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC      or send an email to license@isb-sib.ch).
CC CC      -----
CC CC      EMBL; U39761; G1079712; -.
CC CC      PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC CC      PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
CC CC      PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
CC CC      PFAM; PF00047; ig; 3.
CC CC      PFAM; PF00069; kinase; 1.
CC CC      HSP; P11362; IFGI.
CC CC      KW RECEPTOR; GLYCOPROTEIN; TYROSINE-PROTEIN KINASE; ATP-BINDING;
CC CC      TRANSFERASE; PHOSPHORYLATION; TRANSMEMBRANE; SIGNAL;
CC CC      IMMUNOGLOBULIN FOLD.
CC CC      FT SIGNAL 1 19
CC CC      FT CHAIN 20 1040
CC CC      FT DOMAIN 20 525
CC CC      FT TRANSMEM 526 549
CC CC      FT DOMAIN 550 1040
CC CC      FT DOMAIN 48 116
CC CC      FT DOMAIN 307 374
CC CC      FT DOMAIN 407 492
CC CC      FT DOMAIN 640 931
CC CC      FT NP_BIND 646 654
CC CC      FT BINDING 672 672
CC CC      FT ACT_SITE 797 797
CC CC      FT MOD_RES 828 828
CC CC      FT DISULFID 314 367
CC CC      FT DISULFID 414 485
CC CC      FT CARBOHYD 121 121
CC CC      FT CARBOHYD 280 280
CC CC      FT CARBOHYD 299 299
CC CC      FT CARBOHYD 401 401
CC CC      FT CARBOHYD 407 407
CC CC      FT CARBOHYD 433 433
CC CC      FT CARBOHYD 440 440
CC CC      FT CARBOHYD 474 474
CC CC      FT CARBOHYD 497 497
CC CC      FT MUTAGEN 680 680
CC CC      FT MUTAGEN 714 714
CC CC      FT MUTAGEN 930 930
CC CC      SQ SEQUENCE 1040 AA; 118956 MW; 97C19793 CRC32;

Query Match      67.3%; Score 70; DB 1; Length 1040;
Best Local Similarity 66.7%; Pred.No. 2.74e-02;
Matches      8; Conservative      1; Mismatches      3; Indels      0; Gaps      0;

Db      651 FGEVWKATYKET 662
QY      2 FLAVWKITYKDT 13

RESULT      9
ID      POLG_CXA21      STANDARD;      PRT; 2206 AA.
AC      P22055;
DT      01-AUG-1991 (REL. 19, CREATED)
DT      01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)
DE      15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE      GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS
DE      P2A TO P2C, P3A; GENOME-LINKED PROTEIN VPG; PICORNAIN 3C
DE      (EC 3.4.22.28) (PROTEASE 3C) (P3C); RNA-DIRECTED RNA POLYMERASE P3D
DE      (EC 2.7.7.48)].
OS      COXSACKIEVIRUS A21 (STRAIN COE).
OC      VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;
OC      ENTEROVIRUS.
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE; 90063544.

```


DT 01-JAN-1990 (REL. 13, CREATED)
 DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
 DE 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE MYOSIN HEAVY CHAIN, SKELETAL MUSCLE, ADULT.
 OS GALLUS GALLUS (CHICKEN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
 OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
 RN [1]
 RP SEQUENCE OF 1-205.
 RC TISSUE=PECTORALIS MUSCLE;
 RA MEDLINE; 92041767.
 RA HAYASHIDA M., MAITA T., MATSUDA G.;
 RT "The primary structure of skeletal muscle myosin heavy chain: I.
 RT Sequence of the amino-terminal 23 kDa fragment.";
 RL J. BIOCHEM. 110:54-59(1991).
 RN [2]
 RP SEQUENCE OF 206-636.
 RC TISSUE=PECTORALIS MUSCLE;
 RA MEDLINE; 92041768.
 RA KOMINE Y., MAITA T., MATSUDA G.;
 RT "The primary structure of skeletal muscle myosin heavy chain: II.
 RT Sequence of the 50 kDa fragment of subfragment-1.";
 RL J. BIOCHEM. 110:60-67(1991).
 RN [3]
 RP SEQUENCE OF 637-837.
 RC TISSUE=PECTORALIS MUSCLE;
 RA MEDLINE; 92041769.
 RA MAITA T., MIYANISHI T., MATSUZONO K., TANIOKA Y., MATSUDA G.;
 RT "The primary structure of skeletal muscle myosin heavy chain: III.
 RT Sequence of the 22 kDa fragment and the alignment of the 23 kDa, 50
 RT kDa, and 22 kDa fragments.";
 RL J. BIOCHEM. 110:68-74(1991).
 RN [4]
 RP SEQUENCE OF 838-1938.
 RC TISSUE=PECTORALIS MUSCLE;
 RA MEDLINE; 92041770.
 RA MAITA T., YAJIMA E., NAGATA S., MIYANISHI T., NAKAYAMA S., MATSUDA G.;
 RT "The primary structure of skeletal muscle myosin heavy chain: IV.
 RT Sequence of the rod, and the complete 1,938-residue sequence of the
 RT heavy chain.";
 RL J. BIOCHEM. 110:75-87(1991).
 RN [5]
 RP PRELIMINARY SEQUENCE OF 1-808.
 RX MEDLINE; 87092420.
 RA MAITA T., HAYASHIDA M., TANIOKA Y., KOMINE Y., MATSUDA G.;
 RT "The primary structure of the myosin head.";
 RL PROC. NATL. ACAD. SCI. U.S.A. 84:416-420(1987).
 RN [6]
 RP SEQUENCE OF 842-1270.
 RX MEDLINE; 90121764.
 RA WATANABE B.;
 RT "Complete amino-acid sequence of subfragment-2 in adult chicken
 RT skeletal muscle myosin.";
 RL BIOL. CHEM. HOPPE-SEYLER 370:1027-1034(1989).
 RN [7]
 RP SEQUENCE OF 852-1108.
 RX MEDLINE; 89374803.
 RA WATANABE B.;
 RT "Amino-acid sequence of the short subfragment-2 in adult chicken
 RT skeletal muscle myosin.";
 RL BIOL. CHEM. HOPPE-SEYLER 370:549-558(1989).
 RN [8]
 RP SEQUENCE OF 1145-1270.
 RX MEDLINE; 89228549.
 RA WATANABE B.;
 RT "Amino-acid sequence of the hinge region in chicken myosin
 RT subfragment-2.";
 RL BIOL. CHEM. HOPPE-SEYLER 370:55-61(1989).
 RN [9]
 RP SEQUENCE OF 1857-1938 FROM N.A.
 RX MEDLINE; 87217964.
 RA MORIARTY D.M., BARRINGER K.J., DODGSON J.B., RICHTER H.E.,
 RA YOUNG R.B.;

*Genomic clones encoding chicken myosin heavy-chain genes.";
 DNA 6:91-99(1987).
 RN [10]

RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 1-843.
 RX MEDLINE; 93303624.

RA RAYMENT I., RYPMINSKI W.R., SCHMIDT-BASE K., SMITH R.,
 RA TOMCHICK D.R., BENNING M.M., WINKELMANN D.A., WESENBERG G.,
 RA HOLDEN H.M.;

"Three-dimensional structure of myosin subfragment-1: a molecular
 motor.";
 RT SCIENCE 261:50-58(1993).

CC -1- FUNCTION: MUSCLE CONTRACTION. MYOSIN IS A PROTEIN THAT BINDS TO
 CC F-ACTIN AND HAS ATPASE ACTIVITY THAT IS ACTIVATED BY F-ACTIN.

CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
 CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
 CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).

CC -1- SUBCELLULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.

CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.

CC -1- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
 CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.

CC -1- EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT MEROMYOSIN (LMM)
 CC AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE SPLIT FURTHER INTO
 CC 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED SUBFRAGMENT (S2).

CC -1- SIMILARITY: THE PERIODICITIES OF HYDROPHOBIC AND CHARGED RESIDUES,
 CC WHICH DICTATE THE ALPHA-HELICAL COILED-COIL STRUCTURE ARE
 CC CONSERVED.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----

DR EMBL; M16557; G212372; -

DR PIR; PX0050; PX0051.

DR PIR; A28821; A28821.

DR PIR; S02082; S02082.

DR PIR; S04501; S04501.

DR PIR; S05515; S05515.

DR PDB; 2MYS; 11-JAN-97.

DR PFAM; PF00063; myosin_head; 1.

DR PFAM; PF00612; IQ; 1.

DR KW MYOSIN; MUSCLE PROTEIN; COILED COIL; THICK FILAMENT; ACTIN-BINDING;

DR ATP-BINDING; METHYLATION; ALKYLATION; PHOSPHORYLATION; ACETYLATION;
 DR HEPTAD REPEAT PATTERN; MULTIGENE FAMILY; 3D-STRUCTURE.

FT DOMAIN 1 837 GLOBULAR HEAD (S1).

FT DOMAIN 838 840 HINGE.

FT DOMAIN 841 1880 RODLIKE TAIL (S2 AND LMM DOMAINS).

FT DOMAIN 848 1289 ALPHA-HELICAL TAILPIECE (S2).

FT DOMAIN 1290 1303 HINGE.

FT DOMAIN 1304 1880 LIGHT MEROMYOSIN (LMM).

FT NP_BIND 179 186 COILED COIL (POTENTIAL).

FT DOMAIN 657 679 ATP (POTENTIAL).

FT DOMAIN 759 773 ACTIN-BINDING.

FT MOD_RES 1 1 ACETYLATION.

FT MOD_RES 35 35 METHYLATION (MONO-).

FT MOD_RES 130 130 METHYLATION (TRI-).

FT MOD_RES 551 551 METHYLATION (TRI-).

FT MOD_RES 755 755 METHYLATION (MONO-).

FT MOD_RES 697 697 ALKYLATION (SH-1).

FT MOD_RES 707 707 ALKYLATION (SH-2).

FT CONFLICT 907 907 C -> Q (IN REF. 6 AND 7).

FT CONFLICT 1863 1863 I -> V (IN REF. 9).

FT CONFLICT 1929 1931 IHG -> FH (IN REF. 9).

FT SEQUENCE 1938 AA; 222972 MW; 2ACE77FE CRC32;

Query Match

Score 59; DB 1; Length 1938;

Best Local Similarity 53.8%; Pred. No. 3.21e+00;

```
Matches      7;  Conservative      2;  Mismatches      4;  Indels      0;  Gaps      0;

Db 1453 KILAEWKKYDET 1465
    ||| ||| ||| |||
QY 1 KFLAVWKITYKDI 13

RESULT 13
ID YX09_CAEEL STANDARD; PRT; 230 AA.
AC Q11115;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHEICAL 26.5 KD PROTEIN C03B1.9 IN CHROMOSOME X.
GN C03B1.9.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA MARTIN J.;
RL SUBMITTED (NOV-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U40852; G1072039; -.
DR WORMPEP; C03B1.9; C030910.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 230 AA; 26473 MW; 0F82E358 CRC32;

Query Match      55.8%; Score 58; DB 1; Length 230;
Best Local Similarity 50.0%; Pred. No. 4.81e+00;
Matches      5;  Conservative      2;  Mismatches      3;  Indels      0;  Gaps      0;

Db 74 KFLTKWRTVY 83
    ||| | |
QY 1 KFLAVWKITY 10

RESULT 14
ID SYA_HELPY STANDARD; PRT; 847 AA.
AC P56452;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE ALANYL-TRNA SYNTHETASE (EC 6.1.1.7) (ALANINE--TRNA LIGASE) (ALARS).
GN ALAS OR HPI241.
OS HELICOBACTER PYLORI (CAMPYLOBACTER PYLORI).
OC BACTERIA; PROTEOBACTERIA; EPSILON SUBDIVISION; HELICOBACTER GROUP;
OC HELICOBACTER.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE; 97394467.
RA TOMB J.-F., WHITE O., KERIYAGE A.R., CLAYTON R.A., SUTTON G.G.,
RA FLEISCHMANN R.D., KETCHUM K.A., KLENK H.-P., GILL S., DOUGHERTY B.A.,
RA NELSON K., QUACKENBUSH J., ZHOU L., KIRKNESS E.F., PETERSON S.,
RA LOFTUS B., RICHARDSON D., DODSON R., KHALAK H.G., GLODEK A.,
RA MCKENNEY K., FITZGERALD L.M., LEE N., ADAMS M.D., HICKEY E.K.,
RA BERG D.E., GOCAYNE J.D., UTTERBACK T.R., PETERSON J.D., KELLEY J.M.,
RA COTTON M.D., WEIDMAN J.M., FUJII C., BOWMAN C., WATTHEY L., WALLIN E.,
RA HAYES W.S., BORODOVSKY M., KARP P.D., SMITH H.O., FRASER C.M.,
RA VENTER J.C.;
RA "the complete genome sequence of the gastric pathogen Helicobacter
RT pylori.";
RL NATURE 388:539-547(1997).
```

```
CC -! CATALYTIC ACTIVITY: ATP + L-ALANINE + TRNA(ALA) = AMP +
CC PYROPHOSPHATE + L-ALANYL-TRNA(ALA).
CC -! COFACTOR: BINDS A ZINC ION (BY SIMILARITY).
CC -! SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -! SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -! SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000629; G2314404; -.
DR TIGR; HPI241; -.
DR PROSITE; PS00179; AA_TRNA_LIGASE_IL1; FALSE_NEG.
DR PROSITE; PS00339; AA_TRNA_LIGASE_IL2; 1.
KW AMINOACYL-TRNA SYNTHETASE; PROTEIN BIOSYNTHESIS; LIGASE; ATP-BINDING.
SQ SEQUENCE 847 AA; 94700 MW; 3C3AAD24 CRC32;

Query Match      55.8%; Score 58; DB 1; Length 847;
Best Local Similarity 25.0%; Pred. No. 4.81e+00;
Matches      3;  Conservative      7;  Mismatches      2;  Indels      0;  Gaps      0;

Db 191 RELEIWNLVFMQ 202
    ||| | | | | |
QY 1 KFLAVWKITYKD 12

RESULT 15
ID POLI_HUMAN STANDARD; PRT; 874 AA.
AC P10266;
DT 01-MAR-1989 (REL. 10, CREATED)
DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE RETROVIRUS-RELATED POLYPROTEIN [CONTAINS: REVERSE TRANSCRIPTASE
DE (EC 2.7.7.49); ENDONUCLEASE].
GN POL.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 87036922.
RA ONO M., YASUNAGA T., MIYATA T., USHIKUBO H.;
RT "Nucleotide sequence of human endogenous retrovirus genome related to
RL J. VIROL. 60:589-598(1986).
DR PIR; D24483; GNHUR.
DR PFAM; PF00075; rnaaseh; 1.
DR PFAM; PF00078; rvt; 1.
DR PFAM; PF00552; integrase; 1.
DR PFAM; PF00665; rve; 1.
DR HSSP; P03366; LHMV.
DR HYDROLASE; TRANSFERASE; RNA-DIRECTED DNA POLYMERASE; NUCLEASE;
KW ENDONUCLEASE; POLYPROTEIN.
FT CHAIN 36 250 REVERSE TRANSCRIPTASE.
FT CHAIN 585 764 ENDONUCLEASE.
SQ SEQUENCE 874 AA; 98935 MW; ADFE3749 CRC32;

Query Match      55.8%; Score 58; DB 1; Length 874;
Best Local Similarity 66.7%; Pred. No. 4.81e+00;
Matches      6;  Conservative      2;  Mismatches      1;  Indels      0;  Gaps      0;

Db 722 KFLSQWKS 730
    ||| | | |
QY 1 KFLAVWKIT 9

Search completed: Tue Aug 17 16:03:36 1999
Job time : 12 secs.
```

Tue Aug 17 16:08:17 1999

US-09-049-847-4.rsp

Page 11

This Page Blank (uspto)

M P S R C H

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd
Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Tue Aug 17 16:03:53 1999; MasPar time 5.98 Seconds
Tabular output not generated.
118.713 Million cell updates/sec
Title: >US-09-049-847-4
Description: (1-13) from US09049847.pep
Perfect Score: 104
Sequence: 1 KFLAVWKITYKDT 13
Scoring table: PAM 150
Gap 15
Searched: 179066 seqs, 54579741 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries
Database: sptremb19
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phase 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus
Statistics: Mean 26.892; Variance 39.138; scale 0.687
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	87	83.7	302	14	POLYPROTEIN (FRAGMENT)	2.19e-05
2	87	83.7	302	14	GENOME OF HUMAN POLIOV	2.19e-05
3	87	83.7	879	14	GENOME OF HUMAN POLIOV	2.19e-05
4	87	83.7	2221	14	GENOME OF HUMAN POLIOV	2.19e-05
5	85	81.7	100	14	VP1 PROTEIN (FRAGMENT)	5.81e-05
6	85	81.7	302	14	GENOME OF HUMAN POLIOV	5.81e-05
7	85	81.7	912	14	POLYPROTEIN (FRAGMENT)	5.81e-05
8	82	78.8	2207	14	PV2 POLYPROTEIN.	2.47e-04
9	79	76.0	300	14	(LEON) FRAGMENT ENCODI	1.03e-03
10	79	76.0	300	14	VP1 (FRAGMENT)	1.03e-03
11	79	76.0	300	14	POLIOVIRUS P3/LEON/37	1.03e-03
12	79	76.0	1628	14	COMPLETE SEQUENCE (STR	1.03e-03
13	79	76.0	2206	14	PV3 POLYPROTEIN.	1.03e-03
14	77	74.0	878	14	CAPSID PROTEINS (FRAGM	2.63e-03
15	77	74.0	878	14	CAPSID PROTEINS (FRAGM	2.63e-03
16	77	74.0	878	14	CAPSID PROTEINS (FRAGM	2.63e-03
17	77	74.0	878	14	CAPSID PROTEINS (FRAGM	2.63e-03
18	74	71.2	878	14	CAPSID PROTEINS (FRAGM	1.05e-02
19	70	67.3	1103	5	EG1-15 PROTEIN.	6.39e-02
20	67	54.4	250	1	MOLYBDENUM COPACITOR BI	2.39e-01

21	61	58.7	255	1	Q59026	HYPOTHEICAL PROTEIN M	3.03e+00
22	61	58.7	366	5	Q00843	CAMP-DEPENDENT PROTEIN	3.03e+00
23	61	58.7	500	2	O87172	AMYLALFALASE.	3.03e+00
24	61	58.7	642	14	Q68404	ORF UL150.	3.03e+00
25	61	58.7	940	10	Q81502	F9D12.2 PROTEIN.	3.03e+00
26	60	57.7	261	5	Q21543	COSMID R02E12.	4.55e+00
27	60	57.7	452	5	O17816	F15A4.1 PROTEIN.	4.55e+00
28	60	57.7	585	4	O75539	HYPOTHEICAL 63.6 KD P	4.55e+00
29	59	56.7	345	2	O83918	LIPASE, PUTATIVE.	6.82e+00
30	59	56.7	426	5	Q18083	SIMILARITY TO ARTEMIA	6.82e+00
31	59	56.7	882	11	Q63939	MYOSIN HEAVY CHAIN.	6.82e+00
32	59	56.7	1939	13	O13228	MYOSIN HEAVY CHAIN.	6.82e+00
33	58	55.8	740	14	P87890	POL PROTEIN (FRAGMENT)	1.02e+01
34	58	55.8	1361	4	Q14273	POL/ENV ORF.	1.02e+01
35	58	55.8	1938	6	Q28641	MYOSIN HEAVY CHAIN.	1.02e+01
36	58	55.8	2017	14	Q88285	GAG-POL POLYPROTEIN.	1.02e+01
37	57	54.8	138	2	Q51051	PARIETAL LECTIN (FRAGM	1.51e+01
38	57	54.8	373	8	Q35905	VARI1P.	1.51e+01
39	57	54.8	404	8	O95946	MITOCHONDRION TRANSFER	1.51e+01
40	57	54.8	435	5	O61874	ZK6.10 PROTEIN.	1.51e+01
41	57	54.8	555	10	O80555	T22J18.17 PROTEIN.	1.51e+01
42	57	54.8	599	3	Q99257	CHROMOSOME XVI READING	1.51e+01
43	57	54.8	633	5	O62227	F35E2.7 PROTEIN.	1.51e+01
44	57	54.8	644	14	Q65732	BLUETONGUE VIRUS (17M1	1.51e+01
45	57	54.8	644	14	Q65751	LEUCINE ZIPPER PROTEIN	1.51e+01

ALIGNMENTS

RESULT 1
ID O92270 PRELIMINARY; PRT; 302 AA.
AC O92270;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE POLYPROTEIN (FRAGMENT).
OS HUMAN POLIOVIRUS 1.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;
OC ENTEROVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DAY 11 ISOLATE;
RX MEDLINE: 98411400.
RA KEW O.M., SUTTER R.W., NOTTAY B.K., MCDONOUGH M.J., PREVOIS D.R.,
RA QUICK L., PALLANESCH M.A.;
RT Prolonged replication of a type 1 vaccine-derived poliovirus in an
immunodeficient patient.;
RL J. CLIN. MICROBIOL. 36:2893-2899(1998).
DR EMBL; AF083931; G3617982; -
KW POLYPROTEIN.
FT NON_TER 1 1
FT NON_TER 302 302
SQ SEQUENCE 302 AA; 33505 MW; D99FCE00 CRC32;
Query Match 83.7%; Score 87; DB 14; Length 302;
Best Local Similarity 84.6%; Pred. No. 2.19e-05;
Matches 11; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 103 KLFVWKITYKDT 115
QY 1 KFLVWKITYKDT 13
(::|||||)
RESULT 2
ID Q84871 PRELIMINARY; PRT; 302 AA.
AC Q84871;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE GENOME OF HUMAN POLIOVIRUS TYPE 1 (MAHONEY STRAIN).
DE (ONE OF TWO VERSIONS.).
OS POLIOVIRUS TYPE 1.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;

```

OC ENTEROVIRUS.
RN
RP SEQUENCE FROM N.A.
RC STRAIN=MAHONEY STRAIN;
RX MEDLINE; 81220953.
RA KITAMURA N., SEMLER B.L., ROTHBERG P.G., LARSEN G.R., ADLER C.J.,
RA DORNER A.J., EMINI E.A., HANECAK R., LEE J.J., DER WERF S.,
RA ANDERSON C.W., WIMMER E.;
RT "Primary structure, gene organization and polypeptide expression of
RT poliovirus RNA."
RL NATURE 291:547-553(1981).
RN
RP SEQUENCE FROM N.A.
RC STRAIN=MAHONEY STRAIN;
RX MEDLINE; 82060159.
RA RACANELLO V.R., BALTIMORE D.;
RT "Molecular cloning of poliovirus cDNA and determination of the
RT complete nucleotide sequence of the viral genome."
RL PROC. NATL. ACAD. SCI. U.S.A. 78:4887-4891(1981).
RN
RP SEQUENCE FROM N.A.
RC STRAIN=MAHONEY STRAIN;
RX MEDLINE; 82216986.
RA EMINI E.A., ELZINGA M., WIMMER E.;
RT "Carboxy-terminal analysis of poliovirus proteins: termination of
RT poliovirus RNA translation and location of unique poliovirus
RT polypeptide cleavage sites."
RL J. VIROL. 42:194-199(1982).
RN
RP SEQUENCE FROM N.A.
RC STRAIN=MAHONEY STRAIN;
RX MEDLINE; 82242310.
RA DORNER A.J., DORNER L.F., LARSEN G.R., WIMMER E., ANDERSON C.W.;
RT "Identification of the initiation site of poliovirus polypeptide
RT synthesis."
RL J. VIROL. 42:1017-1028(1982).
DR EMEL; V01148; E9201; -.
DR PFAM; PF00073; rhv; 1.
SQ SEQUENCE 302 AA; 33523 MW; C199C130 CRC32;
Query Match 83.7%; Score 87; DB 14; Length 302;
Best Local Similarity 84.6%; Pred. No. 2.19e-05;
Matches 11; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 103 KLFVAVWKITYKDT 115
QY 1 KFLAVWKITYKDT 13

RESULT 3
ID Q84866 PRELIMINARY; PRT; 879 AA.
AC Q84866;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE GENOME OF HUMAN POLIOVIRUS TYPE 1 (MAHONEY STRAIN).
DE (ONE OF TWO VERSIONS.).
OS POLIOVIRUS TYPE 1.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;
OC ENTEROVIRUS.
RN [1]
RP SEQUENCE OF 1993-2221 FROM N.A.
RC STRAIN=MAHONEY STRAIN;
RX MEDLINE; 81013850.
RA KITAMURA N., WIMMER E.;
RT "Sequence of 1060 3'-terminal nucleotides of poliovirus RNA as
RT determined by a modification of the dideoxynucleotide method."
RL PROC. NATL. ACAD. SCI. U.S.A. 77:3196-3200(1980).
RN [2]
RP SEQUENCE OF 1553-1609 FROM N.A.
RC STRAIN=MAHONEY STRAIN;
RX MEDLINE; 81001866.
RA KITAMURA N., ADLER C.J., ROTHBERG P.G., MARTINKO J., NATHENSON S.G.,
RA WIMMER E.;
RT "The genome-linked protein of picornaviruses. VII. Genetic mapping of
RT poliovirus VPg by protein and RNA sequence studies."
RL CELL 21:285-302(1980).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=MAHONEY STRAIN;
RX MEDLINE; 81220953.
RA KITAMURA N., SEMLER B.L., ROTHBERG P.G., LARSEN G.R., ADLER C.J.,
RA DORNER A.J., EMINI E.A., HANECAK R., LEE J.J., DER WERF S.,
RA ANDERSON C.W., WIMMER E.;
RT "Primary structure, gene organization and polypeptide expression of
RT poliovirus RNA."
RL NATURE 291:547-553(1981).
RN
RP SEQUENCE FROM N.A.
RC STRAIN=MAHONEY STRAIN;

```

```

RX MEDLINE; 82060159.
RA RACANELLO V.R., BALTIMORE D.;
RT "Molecular cloning of poliovirus cDNA and determination of the
RT complete nucleotide sequence of the viral genome."
RL PROC. NATL. ACAD. SCI. U.S.A. 78:4887-4891(1981).
RN
RP SEQUENCE FROM N.A.
RC STRAIN=MAHONEY STRAIN;
RX MEDLINE; 82216986.
RA EMINI E.A., ELZINGA M., WIMMER E.;
RT "Carboxy-terminal analysis of poliovirus proteins: termination of
RT poliovirus RNA translation and location of unique poliovirus
RT polypeptide cleavage sites."
RL J. VIROL. 42:194-199(1982).
RN
RP SEQUENCE FROM N.A.
RC STRAIN=MAHONEY STRAIN;
RX MEDLINE; 82242310.
RA DORNER A.J., DORNER L.F., LARSEN G.R., WIMMER E., ANDERSON C.W.;
RT "Identification of the initiation site of poliovirus polypeptide
RT synthesis."
RL J. VIROL. 42:1017-1028(1982).
DR EMEL; V01148; E9174; -.
DR PFAM; PF00073; rhv; 3.
SQ SEQUENCE 879 AA; 97251 MW; 9F584E23 CRC32;
Query Match 83.7%; Score 87; DB 14; Length 879;
Best Local Similarity 84.6%; Pred. No. 2.19e-05;
Matches 11; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 680 KLFVAVWKITYKDT 692
QY 1 KFLAVWKITYKDT 13

RESULT 4
ID Q84865 PRELIMINARY; PRT; 2221 AA.
AC Q84865;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE GENOME OF HUMAN POLIOVIRUS TYPE 1 (MAHONEY STRAIN)
DE (ONE OF TWO VERSIONS.).
OS POLIOVIRUS TYPE 1.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;
OC ENTEROVIRUS.
RN [1]
RP SEQUENCE OF 1993-2221 FROM N.A.
RC STRAIN=MAHONEY STRAIN;
RX MEDLINE; 81013850.
RA KITAMURA N., WIMMER E.;
RT "Sequence of 1060 3'-terminal nucleotides of poliovirus RNA as
RT determined by a modification of the dideoxynucleotide method."
RL PROC. NATL. ACAD. SCI. U.S.A. 77:3196-3200(1980).
RN [2]
RP SEQUENCE OF 1553-1609 FROM N.A.
RC STRAIN=MAHONEY STRAIN;
RX MEDLINE; 81001866.
RA KITAMURA N., ADLER C.J., ROTHBERG P.G., MARTINKO J., NATHENSON S.G.,
RA WIMMER E.;
RT "The genome-linked protein of picornaviruses. VII. Genetic mapping of
RT poliovirus VPg by protein and RNA sequence studies."
RL CELL 21:285-302(1980).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=MAHONEY STRAIN;
RX MEDLINE; 81220953.
RA KITAMURA N., SEMLER B.L., ROTHBERG P.G., LARSEN G.R., ADLER C.J.,
RA DORNER A.J., EMINI E.A., HANECAK R., LEE J.J., DER WERF S.,
RA ANDERSON C.W., WIMMER E.;
RT "Primary structure, gene organization and polypeptide expression of
RT poliovirus RNA."
RL NATURE 291:547-553(1981).
RN
RP SEQUENCE FROM N.A.
RC STRAIN=MAHONEY STRAIN;

```

RN [4]
 RC SEQUENCE FROM N.A.
 RP STRAIN=MAHONEY STRAIN;
 RX MEDLINE; 82050159.
 RA RACANELLO V.R., BALTIMORE D.;
 RT "Molecular cloning of poliovirus cDNA and determination of the
 complete nucleotide sequence of the viral genome.";
 RL PROC. NATL. ACAD. SCI. U.S.A. 78:4887-4891(1981).
 RN [5]
 RC SEQUENCE FROM N.A.
 RP STRAIN=MAHONEY STRAIN;
 RX MEDLINE; 82216986.
 RA EMINI E.A., ELZINGA M., WIMMER E.;
 RT "Carboxy-terminal analysis of poliovirus proteins: termination of
 poliovirus RNA translation and location of unique poliovirus
 polyprotein cleavage sites.";
 RL J. VIROL. 42:194-199(1982).
 RN [6]
 RC SEQUENCE FROM N.A.
 RP STRAIN=MAHONEY STRAIN;
 RX MEDLINE; 82242310.
 RA DORNER A.J., DORNER L.F., LARSEN G.R., WIMMER E., ANDERSON C.W.;
 RT "Identification of the initiation site of poliovirus polyprotein
 synthesis.";
 RL J. VIROL. 42:1017-1028(1982).
 DR EMBL; Y01148; E9172; -.
 DR PFAM; PF00073; rhv; 3.
 DR PFAM; PF00548; Cys-protease-3C; 1.
 DR PFAM; PF00680; RNA_dep_RNA_pol; 1.
 DR PFAM; PF00910; RNA_helicase; 1.
 DR PFAM; PF00947; Pico_P2A; 1.
 SQ SEQUENCE 2221 AA; 24205 MW; 448A672 CRC32;

Query Match 83.7%; Score 87; DB 14; Length 2221;
 Best Local Similarity 84.6%; Pred. No. 2.19e-05;
 Matches 11; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 695 KLFVAVKITYKDT 707
 QY 1 KFLAVWKITYKDT 13
 |:::|||||

RESULT 5
 ID O91261 PRELIMINARY; PRT; 100 AA.
 AC O91261;
 DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
 DE VPI PROTEIN (FRAGMENT).
 GN VPI.
 OS HUMAN POLIOVIRUS 1.
 OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;
 CC ENTEROVIRUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=3786ALB96;
 RX MEDLINE; 98312955.
 RA FIORE L., GENOVESE D., DIAMANTI E., CATONE S., RIDOLFI B.,
 RA IBRAHIMI B., KONOMI R., VAN DER AVOORT H.G., HOVI T., CRAINC R.,
 RA SIMEONI P., AMATO C.;
 RT "Antigenic and Molecular Characterization of Wild Type 1 Poliovirus
 Causing Outbreaks of Poliomyelitis in Albania and Neighboring
 Countries in 1996.";
 RL J. CLIN. MICROBIOL. 36:1912-1918(1998).
 DR EMBL; AJ007966; E1311747; -.
 FT NON_TER 1
 FT NON_TER 100
 SQ SEQUENCE 100 AA; 10695 MW; 736C541C CRC32;

Query Match 81.7%; Score 85; DB 14; Length 100;
 Best Local Similarity 76.9%; Pred. No. 5.81e-05;
 Matches 10; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 86 KLFVWKITYKDT 98
 QY 1 KFLAVWKITYKDT 13
 |:::|||||

RESULT 6
 ID O84884 PRELIMINARY; PRT; 302 AA.
 AC O84884;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DE GENOME OF HUMAN POLIOVIRUS, STRAIN SABIN 1.
 OS POLIOVIRUS.
 OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;
 CC ENTEROVIRUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SABIN 1;
 RX MEDLINE; 83299876.
 RA NOMOTO A., OMATA T., TOYODA H., KUGE S., HORIE H., KATAOKA Y.,
 RA GENEA Y., NAKANO Y., IMURA N.;
 RT "Complete nucleotide sequence of the attenuated poliovirus Sabin 1
 strain genome.";
 RL PROC. NATL. ACAD. SCI. U.S.A. 79:5793-5797(1982).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SABIN 1;
 RX MEDLINE; 82216986.
 RA EMINI E.A., ELZINGA M., WIMMER E.;
 RT "Carboxy-terminal analysis of poliovirus proteins: termination of
 poliovirus RNA translation and location of unique poliovirus
 polyprotein cleavage sites.";
 RL J. VIROL. 42:194-199(1982).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SABIN 1;
 RX MEDLINE; 82242310.
 RA DORNER A.J., DORNER L.F., LARSEN G.R., WIMMER E., ANDERSON C.W.;
 RT "Identification of the initiation site of poliovirus polyprotein
 synthesis.";
 RL J. VIROL. 42:1017-1028(1982).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SABIN 1;
 RA NOMOTO A.;
 RL SUBMITTED (APR-1988) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; Y01150; E9218; -.
 DR PFAM; PF00073; rhv; 1.
 SQ SEQUENCE 302 AA; 33454 MW; 20DC41FA CRC32;

Query Match 81.7%; Score 85; DB 14; Length 302;
 Best Local Similarity 76.9%; Pred. No. 5.81e-05;
 Matches 10; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 103 KLFVWKITYKDT 115
 QY 1 KFLAVWKITYKDT 13
 |:::|||||

RESULT 7
 ID O89966 PRELIMINARY; PRT; 912 AA.
 AC O89966;
 DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
 DE POLYPROTEIN (FRAGMENT).
 OS HUMAN POLIOVIRUS 1
 OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;
 CC ENTEROVIRUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=97-30515 (16/97/182);
 RA MULDER M.N., REIMERINK J.H.J., STENVIK M., VAN DER AVOORT H.G.A.M.,

```

RA HOVI T., KOOPMANS M.P.G.;
RT "A Sabin Vaccine-Derived Field Isolate of poliovirus Type 1
RT Displaying Aberrant Phenotypic and Genotypic Features, Including a
RT Deletion in Antigenic Site 1.";
RL SUBMITTED (MAY-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF065158; G3511177; -.
KW POLYPROTEIN.
FT NON_TER 912 912
SQ SEQUENCE 912 AA; 100953 MW; 6C3044FD CRC32;

Query Match 81.7%; Score 85; DB 14; Length 912;
Best Local Similarity 76.9%; Pred. No. 5.81e-05;
Matches 10; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 680 KLFVWKITYKDT 692
QY 1 KFLAVWKITYKDT 13

RESULT 8
ID Q98595 PRELIMINARY; PRT; 2207 AA.
AC Q98595;
DT 01-FEB-1997 (TREMELREL. 02, CREATED)
DT 01-FEB-1997 (TREMELREL. 02, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMELREL. 08, LAST ANNOTATION UPDATE)
DE PV2 POLYPROTEIN.
OS HUMAN POLIOVIRUS 2.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;
OC ENTEROVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 84216300.
RA TOYODA H., KOHARA M., KATAOKA Y., SUGANUMA T., OMATA T., IMURA N.,
RA NOMOTO A.;
RT "Complete nucleotide sequences of all three poliovirus serotype
RT genomes. Implication for genetic relationship, gene function and
RT antigenic determinants.";
RL J. MOL. BIOL. 174:561-585(1984).
RN [2]
RP SEQUENCE FROM N.A.
RA NOMOTO A.;
RL SUBMITTED (APR-1988) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; X00595; E275415; -.
DR PFAM; PF00073; rhv; 3.
DR PFAM; PF00548; Cys-protease-3C; 1.
DR PFAM; PF00880; RNA_dep_RNA_pol; 1.
DR PFAM; PF00910; RNA_helicase; 1.
DR PFAM; PF00947; Pico_P2A; 1.
KW POLYPROTEIN; MEMBRANE.
FT CHAIN 1 879 POTENTIAL.
FT CHAIN 1 69 POTENTIAL.
FT CHAIN 70 340 POTENTIAL.
FT CHAIN 341 578 POTENTIAL.
FT CHAIN 579 879 POTENTIAL.
FT CHAIN 880 1454 POTENTIAL.
FT CHAIN 1029 1454 POTENTIAL.
FT CHAIN 1126 1454 POTENTIAL.
FT CHAIN 1455 2207 POTENTIAL.
FT CHAIN 1541 1562 POTENTIAL.
FT CHAIN 1564 2207 POTENTIAL.
FT CHAIN 1747 2207 POTENTIAL.
SQ SEQUENCE 2207 AA; 245967 MW; 709F83D9 CRC32;

Query Match 78.8%; Score 82; DB 14; Length 2207;
Best Local Similarity 69.2%; Pred. No. 2.47e-04;
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 681 RLFSVWKITYKDT 693
QY 1 KFLAVWKITYKDT 13

RESULT 9

```

```

ID Q84791 PRELIMINARY; PRT; 300 AA.
AC Q84791;
DT 01-NOV-1996 (TREMELREL. 01, CREATED)
DT 01-NOV-1996 (TREMELREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMELREL. 08, LAST ANNOTATION UPDATE)
DE (LEON) FRAGMENT ENCODING VP1 (FRAGMENT).
OS POLIOVIRUS TYPE 3.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;
OC ENTEROVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-LEON;
RX MEDLINE; 83141766.
RA MINOR P.D., SCHILD G.C., BOOTMAN J., EVANS D.M.A., FERGUSON M.,
RA REEVE P., SPITZ M., STANWAY G., CANN A.J., HAUPTMANN R., CLARKE L.D.,
RA MOUNTFORD R.C., ALMOND J.W.;
RT "Location and primary structure of a major antigenic site for
RT poliovirus neutralization.";
RL NATURE 301:674-679(1983).
DR EMBL; V01132; G929810; -.
DR PFAM; PF00073; rhv; 1.
FT NON_TER 1 300
FT NON_TER 300 300
SQ SEQUENCE 300 AA; 33458 MW; 71308C5C CRC32;

Query Match 76.0%; Score 79; DB 14; Length 300;
Best Local Similarity 69.2%; Pred. No. 1.03e-03;
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 101 KLFAMWRITYKDT 113
QY 1 KFLAVWKITYKDT 13

RESULT 10
ID Q84891 PRELIMINARY; PRT; 300 AA.
AC Q84891;
DT 01-NOV-1996 (TREMELREL. 01, CREATED)
DT 01-NOV-1996 (TREMELREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMELREL. 08, LAST ANNOTATION UPDATE)
DE VP1 (FRAGMENT).
OS POLIOVIRUS TYPE 3.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;
OC ENTEROVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 84249506.
RA MINOR P.D., EVANS D.M.A., SCHILD G.C., FERGUSON M., ALMOND J.W.;
RT "Identification of an antigenic site in the neutralization of type 3
RT poliovirus.";
RL REV. INFECT. DIS. 6:516-518(1984).
RN [2]
RP SEQUENCE OF 12-300 FROM N.A.
RX MEDLINE; 84249500.
RA ALMOND J.W., CANN A.J., MINOR P.D., REEVE P., SCHILD G.C.,
RA HAUPTMANN R., STANWAY G.;
RT "Nucleotide sequence from neurovirulent and attenuated strains of
RT type 3 poliovirus.";
RL REV. INFECT. DIS. 6:487-493(1984).
DR EMBL; M37321; G332982; -.
DR PFAM; PF00073; rhv; 1.
FT NON_TER 1 300
FT NON_TER 300 300
SQ SEQUENCE 300 AA; 33458 MW; 71308C5C CRC32;

Query Match 76.0%; Score 79; DB 14; Length 300;
Best Local Similarity 69.2%; Pred. No. 1.03e-03;
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 101 KLFAMWRITYKDT 113
QY 1 KFLAVWKITYKDT 13

RESULT 13

```

```

RESULT 11
ID Q84793 PRELIMINARY; PRT; 300 AA.
AC Q84793;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DI 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE POLIOVIRUS P3/LEON/37 GENOME FRAGMENT ENCODING THE VP1 PROTEIN
DE (FRAGMENT).
OS POLIOVIRUS TYPE 3.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;
OC ENTEROVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 84004370.
RA STANWAY G., CANN A.J., HAUPTMANN R., MOUNTFORD R.C., CLARKE L.D.,
RA REEVE P., MINOR P.D., SCHILD G.C., ALMOND J.W.;
RT "Nucleic acid sequence of the region of the genome encoding capsid
RT protein VP1 of neurovirulent and attenuated type 3 polioviruses."
RL EUR. J. BIOCHEM. 135:529-533(1983).
DR EMBL; V01540; G929811; -.
DR PFAM; PF00073; rhv; 1.
FT NON_TER 1 1
FT NON_TER 300 300
SQ SEQUENCE 300 AA; 33458 MW; 71308C5C CRC32;

Query Match 76.0%; Score 79; DB 14; Length 300;
Best Local Similarity 69.2%; Pred. No. 1.03e-03;
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 101 KLFAMWRITYKDT 113
QY 1 KFLAVWKITYKDI 13

RESULT 12
ID Q84784 PRELIMINARY; PRT; 1628 AA.
AC Q84784;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DI 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE COMPLETE SEQUENCE (STRAIN P3/119).
OS POLIOVIRUS TYPE 3.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;
OC ENTEROVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 65037944.
RA CANN A.J., STANWAY G., HUGHES P.J., MINOR P.D., EVANS D.M.A.,
RA SCHILD G.C., ALMOND J.W.;
RT "Reversion to neurovirulence of the live-attenuated Sabin type 3 oral
RT poliovirus vaccine."
RL NUCLEIC ACIDS RES. 12:7787-7792(1984).
DR EMBL; X01076; E9008; -.
DR PFAM; PF00073; rhv; 1.
DR PFAM; PF00548; Cys-protease-3C; 1.
DR PFAM; PF00680; RNA_dep.RNA_pol; 1.
DR PFAM; PF00910; RNA_helicase; 1.
DR PFAM; PF00947; Pico_P2A; 1.
SQ SEQUENCE 1628 AA; 182445 MW; 53BE6A08 CRC32;

Query Match 76.0%; Score 79; DB 14; Length 1628;
Best Local Similarity 69.2%; Pred. No. 1.03e-03;
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 101 KLFAMWRITYKDT 113
QY 1 KFLAVWKITYKDT 13

RESULT 13

```

```

ID Q84792 PRELIMINARY; PRT; 2206 AA.
AC Q84792;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DI 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE P3 POLYPROTEIN.
OS POLIOVIRUS TYPE 3.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;
OC ENTEROVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 84216300.
RA TOYODA H., KOHARA M., KATAOKA Y., SUGANUMA T., OMATA T., IMURA N.,
RA NOMOTO A.;
RT "Complete nucleotide sequences of all three poliovirus serotype
RT genomes. Implication for genetic relationship, gene function and
RT antigenic determinants."
RL J. MOL. BIOL. 174:561-585(1984).
DR EMBL; X00596; G61140; -.
DR PFAM; PF00073; rhv; 3.
DR PFAM; PF00548; Cys-protease-3C; 1.
DR PFAM; PF00680; RNA_dep.RNA_pol; 1.
DR PFAM; PF00910; RNA_helicase; 1.
DR PFAM; PF00947; Pico_P2A; 1.
KW POLYPROTEIN.
SQ SEQUENCE 2206 AA; 246295 MW; 64B5536A CRC32;

Query Match 76.0%; Score 79; DB 14; Length 2206;
Best Local Similarity 69.2%; Pred. No. 1.03e-03;
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 679 KLFAMWRITYKDT 691
QY 1 KFLAVWKITYKDT 13

RESULT 14
ID Q84892 PRELIMINARY; PRT; 878 AA.
AC Q84892;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DI 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE CAPSID PROTEIN (FRAGMENTS).
OS POLIOVIRUS TYPE 3.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;
OC ENTEROVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-SAUKETT COP;
RX MEDLINE; 94160574.
RA HUOVILAINEN A., KINNUNEN L., POYRY T., LAAKSONEN L., ROIVAINEN M.,
RA HOVI T.;
RT "Poliovirus type 3/Saukett: antigenic and structural correlates of
RT sequence variation in the capsid proteins."
RL VIROLOGY 199:228-232(1994).
DR EMBL; L23844; G388329; -.
DR PFAM; PF00073; rhv; 3.
FT CHAIN 1 >69 CAPSID PROTEIN.
FT NON_CONS 69 70
FT CHAIN 70 >340 CAPSID PROTEIN.
FT NON_CONS 340 341
FT CHAIN 341 >578 CAPSID PROTEIN.
FT NON_CONS 578 579
FT CHAIN 579 >878 CAPSID PROTEIN.
FT NON_TER 878 878
SQ SEQUENCE 878 AA; 97326 MW; 1474D049 CRC32;

Query Match 74.0%; Score 77; DB 14; Length 878;
Best Local Similarity 69.2%; Pred. No. 2.63e-03;

```

Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 679 KLFATWRITYKDT 691
|::| |::| |::| |::|
QY 1 KFLAVWKITYKDT 13

RESULT 15
ID 084896 PRELIMINARY; PRJ: 878 AA.
AC 084896;
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DI 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DI 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE CAPSID PROTEINS (FRAGMENTIS).
GN VP-1.
OS POLIOVIRUS TYPE 3.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;
OC ENTEROVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SAUKETT H;
RX MEDLINE: 94160574.
RA HUOVILAINEN A., KINUNEN L., POYRY T., LAAKSONEN L., ROIVAINEN M.,
RA HOVI T.;
RT "Poliovirus type 3/Saukett: antigenic and structural correlates of
sequence variation in the capsid proteins.";
RL VIROLOGY 199:228-232(1994).
DR EMBL: L23848; G388337; -.
DR PFAM: PF00073; rhv; 3.
FT CHAIN 1 >69 CAPSID PROTEIN.
FT NON_CONS 69 70
FT CHAIN 70 >340 CAPSID PROTEIN.
FT NON_CONS 340 341
FT CHAIN 341 >578 CAPSID PROTEIN.
FT NON_CONS 578 579
FT CHAIN 579 >878 CAPSID PROTEIN.
FT NON_TER 878 878
SQ SEQUENCE 878 AA; 97326 MW; 421113F9 CRC32;

Query Match 74.0%; Score 77; DB 14; Length 878;
Best Local Similarity 69.2%; Pred. No. 2.63e-03;
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 679 KLFATWRITYKDT 691
|::| |::| |::| |::|
QY 1 KFLAVWKITYKDT 13

Search completed: Tue Aug 17 16:04:19 1999
Job time : 26 secs.

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.

Copyright (c) 1993-1998 University of Edinburgh, U.K.

Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Aug 17 16:02:04 1999; MasPar time 3.70 Seconds

Tabular output not generated. 74.739 Million cell updates/sec

Title: >US-09-049-847-4

Description: (1-13) from US09049847.pap

Perfect Score: 104

Sequence: 1 KFLAVWKITYKDT 13

Scoring table: PAM 150

Gap 15

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

a-geneseq35

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
 8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
 14:part14 15:part15 16:part16 17:part17 18:part18
 19:part19 20:part20 21:part21 22:part22 23:part23
 24:part24 25:part25 26:part26 27:part27 28:part28
 29:part29 30:part30 31:part31 32:part32 33:part33
 34:part34 35:part35 36:part36 37:part37 38:part38
 39:part39

Statistics: Mean 19.550; Variance 63.096; scale 0.310

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	104	100.0	13	36	Poliovirus antigen.	6.52e-04
2	87	83.7	13	36	Poliovirus antigen.	5.31e-02
3	87	83.7	106	1	Poliovirus VP-1 capsid	5.31e-02
4	87	83.7	106	4	Sequence of a peptide	5.31e-02
5	87	83.7	2209	5	Sequence encoded by a	5.31e-02
6	79	76.0	2206	4	True type 3 poliovirus	3.97e-01
7	62	59.6	111	34	Streptococcus pneumoniae	2.41e-01
8	61	58.7	500	38	Thermus flavus amyloam	3.04e-01
9	61	58.7	642	19	RCMV Toledo strain UL	3.04e-01
10	59	56.7	315	35	Prostate tumour speci	4.83e-01
11	59	56.7	315	36	Amino acid encoded by	4.83e-01
12	59	56.7	665	32	Human PS112 protein s	4.83e-01
13	58	55.8	876	23	Alanine-tRNA synthetas	6.07e-01
14	57	54.8	151	30	Human heart caveolin	7.62e-01
15	57	54.8	434	20	Maga protein.	7.62e-01
16	57	54.8	434	28	Magnetospirillum sp.	7.62e-01

17	56	53.8	1886	31	W54241	Rattus norvegicus mut	9.56e+01
18	55	52.9	1096	1	P92507	pullulanase protein.	1.20e+02
19	54	51.9	14	30	W53471	P2 predominant PNS my	1.50e+02
20	54	51.9	136	30	W40228	Bovine myelin P2 prot	1.50e+02
21	54	51.9	136	30	W40227	Human myelin P2 prote	1.50e+02
22	53	51.0	253	4	R22323	Marek Disease Virus U	1.87e+02
23	53	51.0	355	21	W13105	Marek's disease virus	1.87e+02
24	53	51.0	355	21	W11475	Marek's disease virus	1.87e+02
25	53	51.0	466	22	W09825	UDP-glucose:thiohydro	1.87e+02
26	53	51.0	634	38	W99888	Antigen from cluster	1.87e+02
27	53	51.0	637	20	W11704	High affinity Na-dep	1.87e+02
28	53	51.0	667	31	P97224	Rat prolactin transport	1.87e+02
29	53	51.0	842	4	P93712	Sequence of infection	1.87e+02
30	52	50.0	391	8	R43273	Peptide encoded by la	2.33e+02
31	52	50.0	559	2	R06518	Brassica microspore-s	2.33e+02
32	52	50.0	559	20	W08380	Brassica napus micros	2.33e+02
33	51	49.0	401	7	R39386	GSEA enzyme.	2.90e+02
34	51	49.0	516	39	W67617	P. chrysogenum phenyl	2.90e+02
35	51	49.0	968	28	W41309	CF-5 pathogen resista	2.90e+02
36	50	48.1	538	20	W10058	Human wild-type bile	3.60e+02
37	50	48.1	745	19	R99257	Human bile salt-stimu	3.60e+02
38	50	48.1	1435	13	R70232	P. faiciparum SAPP.	3.60e+02
39	50	48.1	2723	31	W56448	Fragment HGJ1775 of a	3.60e+02
40	50	48.1	2873	39	W89452	Hepatitis G virus PNF	3.60e+02
41	50	48.1	2873	31	W56441	Fragment HGJ605 of a	3.60e+02
42	50	48.1	2873	18	R90796	HGV-PNF 2161 polyprot	3.60e+02
43	50	48.1	2910	39	W89458	Hepatitis G virus var	3.60e+02
44	50	48.1	2969	31	W56442	Fragment HGJ1737 of a	3.60e+02
45	50	48.1	2969	31	W56443	Fragment HGJ1741 of a	3.60e+02

ALIGNMENTS

RESULT 1

ID W67036 standard; peptide; 13 AA.

AC W67036;

DT 15-DEC-1998 (first entry)

DE Poliovirus antigen.

KW Tetanus toxin; vaccine; antibody; carbohydrate peptide conjugate;

KW dendrimeric poly-L-lysine; epitope; tumour.

OS Poliovirus.

PN W09843677-A1.

PD 08-OCT-1998.

PF 27-MAR-1998; E01922.

PR 27-MAR-1997; US-041726.

PA (INSP) INST PASTEUR.

PI Bay S, Cantacuzene D, Leclerc C, Lo-man R;

DR WPI: 98-557071/47.

PT Carbohydrate peptide conjugate used as vaccine - comprises carrier

PT with dendrimeric poly-L-lysine enabling multiple epitopes to be

PT covalently attached

PS Disclosure; Page 14; 55pp; English.

CC The invention relates to a new carbohydrate peptide conjugate, which

CC comprises a carrier with a dendrimeric poly-L-lysine enabling multiple

CC epitopes to be covalently attached to it. Also claimed are: (1) an

CC antibody purified from biological fluid or cells of organisms

CC administered with the carbohydrate peptide conjugate, and (2) a diagnosis

CC kit comprising antigen-specific antibodies elicited by immunisation with

CC the carbohydrate peptide conjugate. The peptide conjugate, antibody and

CC diagnosis kit are used to provide pharmaceutical compositions and

CC vaccines against tumours. These can be used to support an immune response

CC against viral infections caused by hepatitis virus, HIV or cytomegalo

CC virus. They can be used to enhance immune responses, especially B- and T-

CC cell responses, of humans and animals against bacterial infections. The

CC carbohydrate peptide conjugate stimulates the antibody and T-cell

CC response without stimulating undesired immune responses. The composition

CC is capable of increasing the survival of tumour bearing humans and

CC animals. The present sequence represents a poliovirus antigen which

CC forms part of a carbohydrate peptide conjugate.

SQ Sequence 13 AA;

Query Match

Best Local Similarity 100.0%; Score 104; DB 36; Length 13;

Pred. No. 6.52e-04;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 klfavwkitykdt 13
 |||||
 QY 1 KFLAVWKITYKDT 13

RESULT 2

ID W67037 standard; peptide; 13 AA.
 AC W67037;
 DT 15-DEC-1998 (first entry)
 DE Poliovirus antigen.
 KW Tetanus toxin; vaccine; antibody; carbohydrate peptide conjugate;
 KW dendrimeric poly-lysine; epitope; tumour.
 OS Poliovirus.
 PN W03843677-A1.
 PD 08-OCT-1998.
 PF 27-MAR-1998; E01922.
 PR 27-MAR-1997; US-041726.
 PA (INSP) INST PASTEUR.
 PI Ray S, Cantacuzene D, Leclerc C, Lo-man R;
 DR WPI; 98-557071/47.
 PT Carbohydrate peptide conjugate used as vaccine - comprises carrier
 PT with dendrimeric poly-lysine enabling multiple epitopes to be
 PT covalently attached
 PS Example 5; Page 32; 55pp; English.
 CC The invention relates to a new carbohydrate peptide conjugate, which
 CC comprises a carrier with a dendrimeric poly-lysine enabling multiple
 CC epitopes to be covalently attached to it. Also claimed are: (1) an
 CC antibody purified from biological fluid or cells of organisms
 CC administered with the carbohydrate peptide conjugate, and (2) a diagnosis
 CC kit comprising antigen-specific antibodies elicited by immunisation with
 CC the carbohydrate peptide conjugate. The peptide conjugate, antibody and
 CC diagnosis kit are used to provide pharmaceutical compositions and
 CC vaccines against tumours. These can be used to support an immune response
 CC against viral infections caused by hepatitis virus, HIV or cytomegalo
 CC virus. They can be used to enhance immune responses, especially B- and T-
 CC cell responses, of humans and animals against bacterial infections. The
 CC carbohydrate peptide conjugate stimulates the antibody and T-cell
 CC response without stimulating undesired immune responses. The composition
 CC is capable of increasing the survival of tumour bearing humans and
 CC animals. The present sequence represents a poliovirus antigen which
 CC forms part of a carbohydrate peptide conjugate.
 SQ Sequence 13 AA;

Query Match 83.7%; Score 87; DB 36; Length 13;
 Best Local Similarity 84.6%; Pred. No. 5.31e-02;
 Matches 11; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 1 klfavwkitykdt 13
 |||||
 QY 1 KFLAVWKITYKDT 13

RESULT 3

ID P90493 standard; protein; 106 AA.
 AC P90493;
 DT 1-NOV-1989 (first entry)
 DE Poliovirus VP-1 capsid protein
 KW Poliovirus; VP-1 capsid; antibody; vaccine.
 OS Poliovirus
 PN EP-323861-A.
 PD 12-JUL-1989.
 PF 30-NOV-1983; 102888.
 PR 30-NOV-1982; FR-020115.
 PA (INSP) Institut Pasteur.
 PI Girard M, van der Werf S;
 DR WPI; 89-200752/28.
 DR N-PSDB; N90247.
 PT New DNA encoding new immunogenic peptide(s)
 PT - from poliovirus VP-1 protein peptide conjugates and oligomers,
 PT useful in vaccines and diagnosis.
 PS Disclosure; page 3; 53pp; English.

CC The sequence is that of amino acids 23-128 of poliovirus VP-1
 CC capsid polypeptide (see N90247). The peptide is recognised by
 CC antibody C3. It is useful as a reagent for detecting
 CC anti-poliovirus antibodies, and as a component of a vaccine.
 CC It contains the necessary antigenic determinants, but is small
 CC enough to be synthesised chemically.
 SQ Sequence 106 AA;

Query Match 83.7%; Score 87; DB 1; Length 106;
 Best Local Similarity 84.6%; Pred. No. 5.31e-02;
 Matches 11; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 81 klfavwkitykdt 93
 |||||
 QY 1 KFLAVWKITYKDT 13

RESULT 4

ID P40106 standard; Protein; 106 AA.
 AC P40106;
 DT 31-JAN-1992 (first entry)
 DE Sequence of a peptide recognised by polio antibodies.
 KW Polio virus capsid; VP-1 structural protein; C particle; D particle;
 KW vaccine; diagnosis.
 OS Polio virus.
 PN EP-110791-A.
 PD 13-JUN-1984.
 PF 30-NOV-1983; 402310.
 PR 30-NOV-1982; FR-020115.
 PR 29-JUN-1983; FR-010778.
 PR 01-JAN-1989; EP-102888.
 PA (INSP) INST PASTEUR.
 PA (GIRA/) GIRARD M.
 PI Girard M, Van der Werf S;
 DR WPI; 84-147943/24.
 DR N-PSDB; N40084.
 PT DNA fragment coding for peptide recognised by polio antibodies -
 PT its derived peptide(s) and oligomers and transformed
 PT microorganisms
 PS Claim 9; Page 33; 46pp; French.
 CC N40084 codes for a peptide (P40106) which can be recognised by
 CC antibodies active against C and D particles of the same polio virus
 CC and against the VP-1 structural polypeptide of this polio virus
 CC capsid.
 SQ Sequence 106 AA;

Query Match 83.7%; Score 87; DB 4; Length 106;
 Best Local Similarity 84.6%; Pred. No. 5.31e-02;
 Matches 11; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 81 klfavwkitykdt 93
 |||||
 QY 1 KFLAVWKITYKDT 13

RESULT 5

ID P20037 standard; Protein; 2209 AA.
 AC P20037;
 DT 05-AUG-1992 (first entry)
 DE Sequence encoded by a full-length cDNA copy of the poliovirus genome
 DE in plasmid pVRI06.
 KW Poliovirus; picornavirus; vaccine; antigen; immunogen.
 OS Human poliovirus.
 FH Key Location/Qualifiers
 FT protein 1..69
 FT /label= P4
 FT protein 70..341
 FT /label= VP2
 FT protein 342..579
 FT /label= VP3
 FT protein 580..881
 FT /label= VP1
 FT protein 882..1030


```

DT 01-MAR-1999 (first entry)
DE Thermus flavus amyloamaltase.
KW Thermus flavus; amyloamaltase; heat resistant; cyclic glucan;
KW intramolecular transglycosylation; alpha-glucan; food.
OS Thermus flavus.
PN EP-884384-A2.
PD 16-DEC-1998.
PF 13-MAY-1998; 250162.
PR 07-MAY-1998; JP-125121.
PR 13-MAY-1997; JP-122635.
PA (EZAKI) EZAKI GLICO CO LTD.
PI Fujii K, Okada S, Takahata T, Takata H, Terada Y,
PI Yanase M;
DR WPI; 99-026580/03.
DR N-PSDB; V72539.
PT New amylo-maltase catalyses intra-molecular trans-glycosylation of
PT alpha-glucans - used as additive to e.g. rice products, snacks,
PT wheat products, noodles, processed seafood, frozen or refrigerated
PT foods, baby foods or drinks
PS Claim 3; Page 17-18; 32pp; English.
CC The present sequence is an amyloamaltase isolated from Thermus flavus.
CC Amyloamaltase catalyses intramolecular transglycosylation of alpha-glucans
CC to generate cyclic glucans, has no hydrolase activity, has an optimum
CC temperature of 65-70 degrees Celsius, remains active at 60 degrees
CC Celsius for at least 10 minutes, is inactivated at 100 degrees Celsius
CC within 15 minutes and has an optimum pH of 5.5. Amyloamaltase can be used
CC to produce a cyclic glucan by cyclising an alpha-glucan and collecting
CC and purifying the cyclic glucan (especially where the cyclic glucan
CC comprises a cyclic alpha-1,4-glucan, a branched cyclic glucan and/or a
CC branching enzyme is also used in the cyclisation step). The amyloamaltase
CC is used to produce food by adding it to a food material before or
CC immediately after cooking so that the amyloamaltase acts on starch in the
CC food material to produce a cyclic glucan (especially where the food is
CC selected from rice products, Japanese desserts, snacks, wheat products,
CC noodles, gyoza skins, shumai skins, processed seafoods, frozen or
CC refrigerated processed foods, weaning foods, baby foods, pet foods,
CC animal feeds, drinks, sports foods and nutritional supplements).
CC Sequence 500 AA;
SQ Sequence 500 AA;

Query Match 58.7%; Score 61; DB 38; Length 500;
Best Local Similarity 41.7%; Pred. No. 3.04e+01;
Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

DB 416 ryladwgifre 427
::||| ||:::
QY 1 KFLAVKIIYKDT 12

RESULT 9
ID W05519 standard; Protein; 642 AA.
AC W05519;
DT 15-JAN-1997 (first entry)
DE HCMV Toledo strain ULI50 protein (clone tol.21).
KW CMV; HCMV; vaccine; diagnosis; ULI50.
OS Human cytomegalovirus Toledo strain.
PN W09630387-A1.
PD 03-OCT-1996.
PF 26-MAR-1996; U04100.
PR 31-MAR-1995; US-414926.
PA (AVIR-) AVIRON.
PI Cha T, Spaete R;
DR WPI; 96-45265/45.
DR N-PSDB; T41418.
PT New isolated human cytomegalovirus nucleic acid - from Towne and
PT Toledo strains, used to develop prods. for the diagnosis, prevention
PT and treatment of human CMV infections
PS Claim 5; Page 85-88; 150pp; English.
CC Novel protein ULI50 (W05519) is the product of an open reading
CC frame found in a novel nucleic acid (T41418) isolated from the
CC Toledo strain of human cytomegalovirus (HCMV). ULI50 and other
CC novel (see also W05502-20) and known (see also W05500-01) proteins
CC of the Toledo strain, as well as new proteins (see also W05521-24)
CC from HCMV Towne, can be produced in transformed host cells and used

```

```

CC in the prodn. of subunit vaccines against HCMV. They may be
CC surface glycoproteins that are immunogenic or responsible for
CC tissue tropism, or may influence the immune response of an infected
CC individual.
SQ Sequence 642 AA;

Query Match 58.7%; Score 61; DB 19; Length 642;
Best Local Similarity 63.6%; Pred. No. 3.04e+01;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

DB 404 laiwqrtyndt 414
||:| ||::||
QY 3 LAVWKIIYKDT 13

RESULT 10
ID W69384 standard; Protein; 315 AA.
AC W69384;
DT 08-DEC-1998 (first entry)
DE Prostate tumour specific gene clone J1-17 protein.
KW Prostate tumour specific gene; human; prostate cancer; detection;
KW therapy.
OS Homo sapiens.
PN W09837418-A2.
PD 27-AUG-1998.
PR 25-FEB-1998; U03690.
PR 09-FEB-1998; US-904809.
PR 25-FEB-1997; US-806596.
PR 01-AUG-1997; US-904809.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI; 98-480805/41.
DR N-PSDB; V58585.
PT Novel human prostate specific tumour protein and fragments - useful
PT for detecting and treating prostate cancers
PS Example 1; Page 86-87; 141pp; English.
CC This sequence is encoded by a human prostate tumour specific gene, and
CC can be used in the method of the invention. The method is for detecting
CC prostate cancer comprising contacting a biological sample with an agent
CC able to bind an immunogenic portion of a prostate protein (such as
CC this protein sequence). An antibody which binds to an immunogenic
CC portion of the prostate protein, and the method can be used to detect,
CC monitor progression of, or treat prostate cancers. The antibody may
CC also be conjugated to a therapeutic agent for use in therapy of prostate
CC cancers.
SQ Sequence 315 AA;

Query Match 56.7%; Score 59; DB 35; Length 315;
Best Local Similarity 55.6%; Pred. No. 4.83e-01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

DB 37 flqgvlvay 45
||:| ||::||
QY 2 FLAVWKIIY 10

RESULT 11
ID W71868 standard; Protein; 315 AA.
AC W71868;
DT 06-JAN-1999 (first entry)
DE Amino acid encoded by prostate tumour clone J1-17.
KW Prostate; cancer; tumour; vaccine; immunogen; clone.
OS Homo sapiens.
PN W09837093-A2.
PD 27-AUG-1998.
PR 25-FEB-1998; U03492.
PR 09-FEB-1998; US-020956.
PR 25-FEB-1997; US-806099.
PR 01-AUG-1997; US-904804.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI; 98-609886/51.
DR N-PSDB; V61200.

```

PT Polypeptides comprising immunogenic portions of prostate proteins -
 PT used in a vaccine for the treatment of prostate cancer
 PS Example 1; Page 81-82; 130pp; English.
 CC The present sequence is an immunogenic portion of a prostate tumour
 CC protein. The immunogen, or the DNA encoding it, can be used as a
 CC vaccine for the treatment of prostate cancer. The immunogen was
 CC isolated from a prostate tumour cDNA library obtained by subtracting
 CC a prostate tumour cDNA expression library with a normal tissue cDNA
 CC library.
 SQ Sequence 315 AA;

Query Match 56.7%; Score 59; DB 36; Length 315;
 Best Local Similarity 55.6%; Pred. No. 4.83e+01;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 37 flgwvlvay 45
 |||||
 QY 2 FLAVWKITY 10

RESULT 12

ID W54425 standard; Protein; 665 AA.

AC W54425;
 DT 15-SEP-1998 (first entry)
 DE Human PS112 protein sequence from gene-specific clones.
 KW Prostate; disease; PS112 gene; detection; diagnosis; cancer;
 KW treatment; antibody.

OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Protein 1..665

FT /label= PS112

FT /note= "partial sequence"

PN W0981557-A1.

PD 16-APR-1998.
 PE 08-OCT-1997; U18290.
 PR 08-OCT-1996; US-727688.
 PA (ABBO) ABBOTT LAB.
 PI Cohen M. Friedman PN, Gordon J, Hodges SC, Klass MR,
 PI Kratochvil JD, Roberts-Rapp L, Russell JC, Stroupe SD;
 DR WPI; 98-240838/21.
 PT Detecting a target PS112 polynucleotide - used for diagnosing
 PT prostate cancer

PT Example 10; Page 89-90; 104pp; English.

PS This sequence represents the human PS112 protein derived from various
 CC gene specific clones isolated from a prostate library and is used in a
 CC novel method of detecting the presence of a target PS112 polynucleotide
 CC in a test sample. The method can also be used to detect mRNA of PS112 in
 CC a test sample. The method can be used for diagnosis of prostate cancer,
 CC as the presence of PS112 is an indicator of prostate cancer. Antibodies
 CC against the polypeptides may be used as markers, or to treat prostate
 CC cancer.

SQ Sequence 665 AA;

Query Match 55.7%; Score 59; DB 32; Length 665;
 Best Local Similarity 55.6%; Pred. No. 4.83e-01;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 388 flgwvlvay 396
 |||||
 QY 2 FLAVWKITY 10

RESULT 13

ID W21898 standard; Protein; 876 AA.

AC W21898;
 DT 11-SEP-1997 (first entry)
 DE Alanyl-tRNA synthetase from Staph. aureus.
 KW tRNA synthetase; Escherichia coli; immunological response; antibody;
 KW bacterial infection; adherence; damaged tissue; wound healing;
 KW vaccine; skin; protection.
 OS Staphylococcus aureus.

PN EP-785260-A1.

PD 23-JUL-1997.

PF 17-JAN-1997; 300309.
 PR 19-JAN-1996; GB-001099.
 PR 30-OCT-1996; GB-022617.
 PA (SWIK) SMITHKLINE BEECHAM PLC.
 PI Hodgson JE, Lawlor EU;
 DR WPI; 97-365935/34.
 DR N-PSDB; T73696.
 PT DNA encoding alanyl-tRNA synthetase from Staphylococcus aureus WCUH
 PT 29 - useful for protection against bacterial infections
 PS Claim 1; Fig 2; 35pp; English.

CC The present sequence represents a novel alanyl-tRNA synthetase protein
 CC from Staphylococcus aureus strain WCUH29 (NCIMB 40771), which is related
 CC by amino acid sequence homology to Escherichia coli alanyl tRNA
 CC synthetase. Vectors comprising the DNA (or polynucleotides having at
 CC least 70 % identity to it) can be used for the recombinant production
 CC of the enzyme. The enzyme or its related DNA (through gene therapy) is
 CC used to induce an immunological response in a mammal to generate
 CC antibodies to protect against disease. The antibodies protect against
 CC invasion of bacteria, e.g. by blocking adherence of bacteria to damaged
 CC tissue, including wounds in skin or connective tissue caused by
 CC mechanical, chemical or thermal damage or by implantation of in-dwelling
 CC devices, or wounds in the mucous membranes. Antagonists are used to
 CC inhibit the enzyme, especially to prevent adhesion of bacteria to
 CC mammalian extracellular matrix proteins on in-dwelling devices or to
 CC extracellular matrix proteins in wounds, or to block tRNA synthetase
 CC protein mediated mammalian cell invasion by, e.g. initiating
 CC phosphorylation of mammalian tyrosine kinase. Analysing a sample for
 CC the presence of the enzyme (or a polypeptide having at least 70 %
 CC identity to it) is used for a diagnostic process.

SQ Sequence 876 AA;

Query Match 55.8%; Score 58; DB 23; Length 876;
 Best Local Similarity 25.0%; Pred. No. 6.07e+01;
 Matches 3; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

Db 197 rylevwnlvse 208
 :|||:::
 QY 1 KFLAVWKITYKD 12

RESULT 14

ID W40495 standard; Protein; 151 AA.

AC W40495;
 DT 14-JUL-1998 (first entry)
 DE Human heart caveolin protein.
 KW Caveolin; human; heart; prophylaxis; treatment; diabetes; obesity;
 KW cancer; arteriosclerosis; muscular dystrophy; inhibitor.

OS Homo sapiens.

PN J10087698-A.

PD 07-APR-1998.

PF 03-JUL-1997; 177496.

PR 09-JUL-1996; JP-179666.

PA (TAKE) TAKEDA CHEM IND LTD.

DR WPI; 98-267126/24.

DR N-PSDB; V11143; V11147.

PT New human heart caveolin protein and related DNA - useful for
 PT preventing and treating diabetes, obesity, cancer, arteriosclerosis
 PT and muscular dystrophy
 PS Claim 1; Fig 1; 25pp; Japanese.

CC This sequence represents a human heart caveolin protein. This protein
 CC can be used for prophylaxis and treatment of diabetes, obesity, cancer,
 CC arteriosclerosis and muscular dystrophy. Caveolin can also be used for
 CC identifying compounds which can promote or inhibit its activity.

SQ Sequence 151 AA;

Query Match 54.8%; Score 57; DB 30; Length 151;
 Best Local Similarity 55.6%; Pred. No. 7.62e-01;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 54 fdsvkvksy 62
 |::|||:
 QY 2 FLAVWKITY 10

```

RESULT 15
ID W01520 standard; Protein; 434 AA.
AC W01520;
DI 28-FEB-1997 (first entry)
DE MagA protein.
KW magA; magnetic; microbe; protein preparation; organic membrane.
OS Magnetospirillum sp.
FH Key
FT Location/Qualifiers
FT region 7..380
FT /label= hydrophobic_region
PN J08228782-A.
PD 10-SEP-1996.
PF 18-SEP-1995; 263487.
PR 16-SEP-1994; JP-248700.
PA (MAIS/) MATSUNAGA T.
PA (DENK ) TDK CORP.
DR WPI; 96-459147/46.
DR N-PSDB; T46127.
PT magA gene encoding protein bound to organic membrane - covering fine
PT magnetic particles formed in magnetic microbe AMB1, useful for
PT stable prepn. of proteins
PS Claim 2; Page 9-11; 17pp; Japanese.
CC The magA protein binds to an organic membrane covering fine magnetic
CC particles formed in the magnetic microbe AMB-1. The magnetic microbe
CC can be used for the stable prepn. of proteins.
SQ Sequence 434 AA;

Query Match 54.8%; Score 57; DB 20; Length 434;
Best Local Similarity 55.6%; Pred. No. 7.62e+01;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 80 rflevwhta 88
QY :|||!|:
QY 1 KFLAVWKIT 9

Search completed: Tue Aug 17 16:02:29 1999
Job time : 25 secs.

```

M P S R C H

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MSPrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Aug 17 16:04:36 1999; MasPar time 1.57 Seconds
83.835 Million cell updates/sec

Tabular output not generated.

Title: >US-09-049-847-4
Description: (1-13) from US09049847.pep
Perfect Score: 104
Sequence: 1 KFLAVWKITYKDI 13

Scoring table: PAM 150
Gap 15

Searched: 106580 seqs, 10152877 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:PC9_COMB 4:backfiles1

Statistics: Mean 18.363; Variance 62.160; scale 0.295

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match Length DB ID	Description Pred. No.
1	87	83.7	13 2 US-08-397- Sequence 3, Applicatio 3.23e-02
2	87	83.7	13 1 US-08-479- Sequence 5, Applicatio 3.23e-02
3	87	83.7	13 1 US-08-336- Sequence 5, Applicatio 3.23e-02
4	79	76.0	2206 1 US-07-852- Sequence 2, Applicatio 2.28e-01
5	79	76.0	2206 2 US-08-461- Sequence 2, Applicatio 2.28e-01
6	61	58.7	642 1 US-08-414- Sequence 25, Applicati 1.58e+01
7	58	55.8	876 2 US-08-785- Sequence 2, Applicatio 3.11e+01
8	57	54.8	434 2 US-08-529- Sequence 2, Applicatio 3.89e+01
9	54	51.9	132 2 US-08-470- Sequence 2, Applicatio 7.55e+01
10	54	51.9	132 1 US-08-409- Sequence 11, Applicati 7.55e+01
11	53	51.0	25 2 US-08-241- Sequence 11, Applicati 9.40e+01
12	53	51.0	25 1 US-08-468- Sequence 14, Applicati 9.40e+01
13	53	51.0	635 1 US-07-879- Sequence 9, Applicatio 9.40e+01
14	53	51.0	635 2 US-08-753- Sequence 8, Applicatio 9.40e+01
15	53	51.0	667 2 US-08-753- Sequence 8, Applicatio 9.40e+01
16	53	51.0	667 1 US-07-879- Sequence 6, Applicatio 1.17e+02
17	52	50.0	559 1 US-08-030- Sequence 6, Applicatio 1.17e+02
18	52	50.0	723 4 5200183-4 Patent No. 5200183. 1.80e+02
19	50	48.1	317 1 US-07-603- Sequence 30, Applicati 1.80e+02
20	50	48.1	546 2 US-08-445- Sequence 5, Applicatio 1.80e+02
21	50	48.1	546 2 US-08-204- Sequence 5, Applicatio 1.80e+02
22	50	48.1	568 2 US-08-445- Sequence 6, Applicatio 1.80e+02
23	50	48.1	571 4 5200183-17 Patent No. 5200183. 1.80e+02

24	50	48.1	668	2	US-08-204- Sequence 9, Applicatio 1.80e+02
25	50	48.1	668	2	US-08-445- Sequence 9, Applicatio 1.80e+02
26	50	48.1	722	4	5200183-3 Patent No. 5200183. 1.80e+02
27	50	48.1	722	2	US-08-482- Sequence 1, Applicatio 1.80e+02
28	50	48.1	722	2	US-08-204- Sequence 7, Applicatio 1.80e+02
29	50	48.1	722	2	US-08-445- Sequence 7, Applicatio 1.80e+02
30	50	48.1	722	2	US-08-445- Sequence 3, Applicatio 1.80e+02
31	50	48.1	722	2	US-08-204- Sequence 3, Applicatio 1.80e+02
32	50	48.1	742	4	5200183-2 Patent No. 5200183. 1.80e+02
33	50	48.1	742	2	US-08-482- Sequence 2, Applicatio 1.80e+02
34	50	48.1	745	2	US-08-204- Sequence 2, Applicatio 1.80e+02
35	50	48.1	1039	4	5196511-2 Patent No. 5196511. 1.80e+02
36	50	48.1	1435	2	US-08-568- Sequence 4, Applicatio 1.80e+02
37	50	48.1	2873	3	PCT-US95-0 Sequence 15, Applicati 1.80e+02
38	50	48.1	2873	2	US-08-464- Sequence 15, Applicati 1.80e+02
39	50	48.1	2873	2	US-08-485- Sequence 15, Applicati 1.80e+02
40	50	48.1	2873	2	US-08-444- Sequence 15, Applicati 1.80e+02
41	50	48.1	2910	3	PCT-US95-0 Sequence 157, Applicat 1.80e+02
42	50	48.1	2910	2	US-08-461- Sequence 183, Applicat 1.80e+02
43	50	48.1	2910	2	US-08-485- Sequence 183, Applicat 1.80e+02
44	50	48.1	2910	2	US-08-464- Sequence 183, Applicat 1.80e+02
45	50	48.1	2910	2	US-08-444- Sequence 183, Applicat 1.80e+02

ALIGNMENTS

RESULT 1
ID US-08-397-286-3 STANDARD; PRT; 13 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 3, Application US/08397286
XX
CC Sequence 3, Application US/08397286
CC Patent No. 5871747
CC GENERAL INFORMATION:
CC APPLICANT: GENGOUX, CHRISTINE; LeCLERC, CLAUDE
CC TITLE OF INVENTION: ANTIGEN-CARRYING
CC TITLE OF INVENTION: MICROPARTICLES AND THEIR USE IN THE IN THE
CC TITLE OF INVENTION: INDUCTION OF HUMORAL OR CELLULAR RESPONSES
CC NUMBER OF SEQUENCES: 5
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: BIERMAN & MUSERLIAN
CC STREET: 600 THIRD AVENUE
CC CITY: NEW YORK
CC STATE: NEW YORK
CC COUNTRY: U.S.A.
CC ZIP: 10016
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: FLOPPY DISK
CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASC II
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/397,286
CC FILING DATE: 10-MAR-1995
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC PRIOR APPLICATION NUMBER: PCT/FR93/00876
CC FILING DATE: 13-SEP-1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: FR/92/10879
CC FILING DATE: 11-SEP-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: CHARLES A. MUSERLIAN
CC REGISTRATION NUMBER: 19,683
CC REFERENCE/DOCKET NUMBER: 102.162
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 212-661-8000
CC TELEFAX: 212-661-8002
CC INFORMATION FOR SEQ ID NO: 3:

[illegible]


```
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Cseri, Luann
CC REGISTRATION NUMBER: 31,822
CC REFERENCE/DOCKET NUMBER: AVIR-011/OOUS
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 415-494-7622
CC TELEFAX: 415-857-0663
CC INFORMATION FOR SEQ ID NO: 25:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 642 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC IMMEDIATE SOURCE:
CC CLONE: tol.21
CC FEATURE:
CC NAME/KEY: Protein
CC LOCATION: 1..642
CC OTHER INFORMATION: /label= ULJ50
CC SEQUENCE 642 AA: 70848 MW: 2000477 CN;

Query Match 58.7%; Score 61; DB 1; Length 642;
Best Local Similarity 63.8%; Pred. No. 1.58e+01;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 404 LAIWQTYNDT 414
QY 3 LAVWKITYKDT 13

RESULT 7
ID US-08-785-071A-2 STANDARD; PRT; 876 AA.
XX XXXXXX
AC XXXXXX
XX XXXXXX
DE Sequence 2, Application US/08785071A
XX Sequence 2, Application US/08785071A
CC Patent No. 5776750
CC GENERAL INFORMATION:
CC APPLICANT: Hodgson, John
CC APPLICANT: Lawlor, Elizabeth
CC TITLE OF INVENTION: NO. 5776750el tRNA Synthetase
CC NUMBER OF SEQUENCES: 2
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: SmithKline Beecham Corporation
CC STREET: 709 Swedeland Road
CC CITY: King of Prussia
CC STATE: PA
CC COUNTRY: USA
CC ZIP: 19406-0939
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: DOS
CC SOFTWARE: FastSeq for Windows Version 2.0
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/785,071A
CC FILING DATE: 17-JAN-1997
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 9601099.6
CC FILING DATE: 19-JAN-1996
CC APPLICATION NUMBER: 9622617.0
CC FILING DATE: 27-JUL-1996
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Gimm, Edward R
CC REGISTRATION NUMBER: 38,891
CC REFERENCE/DOCKET NUMBER: P31355-6
CC TELECOMMUNICATION INFORMATION:

CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Cseri, Luann
CC REGISTRATION NUMBER: 31,822
CC REFERENCE/DOCKET NUMBER: AVIR-011/OOUS
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 415-494-7622
CC TELEFAX: 415-857-0663
CC INFORMATION FOR SEQ ID NO: 25:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 642 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC IMMEDIATE SOURCE:
CC CLONE: tol.21
CC FEATURE:
CC NAME/KEY: Protein
CC LOCATION: 1..642
CC OTHER INFORMATION: /label= ULJ50
CC SEQUENCE 642 AA: 70848 MW: 2000477 CN;

Query Match 58.7%; Score 61; DB 1; Length 642;
Best Local Similarity 63.8%; Pred. No. 1.58e+01;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 404 LAIWQTYNDT 414
QY 3 LAVWKITYKDT 13

RESULT 7
ID US-08-785-071A-2 STANDARD; PRT; 876 AA.
XX XXXXXX
AC XXXXXX
XX XXXXXX
DE Sequence 2, Application US/08785071A
XX Sequence 2, Application US/08785071A
CC Patent No. 5776750
CC GENERAL INFORMATION:
CC APPLICANT: Hodgson, John
CC APPLICANT: Lawlor, Elizabeth
CC TITLE OF INVENTION: NO. 5776750el tRNA Synthetase
CC NUMBER OF SEQUENCES: 2
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: SmithKline Beecham Corporation
CC STREET: 709 Swedeland Road
CC CITY: King of Prussia
CC STATE: PA
CC COUNTRY: USA
CC ZIP: 19406-0939
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: DOS
CC SOFTWARE: FastSeq for Windows Version 2.0
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/785,071A
CC FILING DATE: 17-JAN-1997
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 9601099.6
CC FILING DATE: 19-JAN-1996
CC APPLICATION NUMBER: 9622617.0
CC FILING DATE: 27-JUL-1996
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Gimm, Edward R
CC REGISTRATION NUMBER: 38,891
CC REFERENCE/DOCKET NUMBER: P31355-6
CC TELECOMMUNICATION INFORMATION:

CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Cseri, Luann
CC REGISTRATION NUMBER: 31,822
CC REFERENCE/DOCKET NUMBER: AVIR-011/OOUS
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 415-494-7622
CC TELEFAX: 415-857-0663
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 876 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 876 AA: 98538 MW: 3870699 CN;

Query Match 55.8%; Score 58; DB 2; Length 876;
Best Local Similarity 25.0%; Pred. No. 3.11e+01;
Matches 3; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

Db 197 RYLEVWNLVFE 208
QY 1 KFLAVWKITYKD 12

RESULT 8
ID US-08-529-600D-2 STANDARD; PRT; 434 AA.
XX XXXXXX
AC XXXXXX
XX XXXXXX
DE Sequence 2, Application US/08529600D
XX Sequence 2, Application US/08529600D
CC Patent No. 5861285
CC GENERAL INFORMATION:
CC APPLICANT: Tadashi MATSUNAGA
CC TITLE OF INVENTION: PROTEIN-BOUND MAGNETIC PARTICLES AND
CC TITLE OF INVENTION: PROCESS OF PRODUCING THE SAME
CC NUMBER OF SEQUENCES: 3
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: PILLSBURY MADISON & SUTRO, L.L.P.
CC STREET: 1100 NEW YORK AVENUE, N.W.
CC CITY: WASHINGTON
CC STATE: D.C.
CC COUNTRY: U.S.A.
CC ZIP: 20005-3918
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Microsoft Word
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/529,600D
CC FILING DATE: 18-SEP-1995
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: JP 6-248700
CC FILING DATE: 16-SEP-1994
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 434 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 434 AA: 46827 MW: 982549 CN;

Query Match 54.8%; Score 57; DB 2; Length 434;
Best Local Similarity 55.8%; Pred. No. 3.89e+01;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 80 RFLVWVWKA 88
QY 1 KFLAVWKIT 9
```

RESULT 9
ID US-08-470-298B-11 STANDARD; PRT; 132 AA.
AC xxxxxx
XX
XX
DT
XX
XX
DE Sequence 11, Application US/08470298B
XX
XX Sequence 11, Application US/08470298B
CC Patent No. 5844081
CC GENERAL INFORMATION:
CC APPLICANT: NI, JIAN
CC APPLICANT: GENTZ, REINER
CC APPLICANT: YU, GUO-LIANG
CC APPLICANT: ROSEN, CRAIG A.
CC TITLE OF INVENTION: CYTOSTATIN I
CC NUMBER OF SEQUENCES: 12
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: HUMAN GENOME SCIENCES, INC.
CC STREET: 9410 KEY WEST AVENUE
CC CITY: ROCKVILLE
CC STATE: MD
CC COUNTRY: US
CC ZIP: 20850
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/470.298B
CC FILING DATE: 06-JUN-1995
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: BROOKES, ALLAN A.
CC REGISTRATION NUMBER: 36,373
CC REFERENCE/DOCKET NUMBER: PF175D1
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 301-309-8504
CC TELEFAX: 301-309-8512
CC INFORMATION FOR SEQ ID NO: 11:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 132 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC ORGANISM: MYELIN P2 (FIGURE 2)
CC ORGANISM: MYELIN P2 (FIGURE 2)
SQ SEQUENCE 132 AA; 14909 MW; 94859 CN;

Query Match 51.9%; Score 54; DB 2; Length 132;
Best Local Similarity 62.5%; Pred. No. 7.55e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 4 KFLGTWKL 11
QY 1 KFLAVWKI 8
||||: ||:

RESULT 10
ID US-08-409-731A-11 STANDARD; PRT; 132 AA.
XX
AC xxxxxx
XX
XX
DI
XX
XX
DE Sequence 11, Application US/08409731A
XX
XX Sequence 11, Application US/08409731A
CC Patent No. 5658758
CC GENERAL INFORMATION:

CC APPLICANT: NI, JIAN
CC APPLICANT: YU, GUO-LIANG
CC APPLICANT: GENTZ, REINER
CC APPLICANT: ROSEN, CRAIG A.
CC TITLE OF INVENTION: CYTOSTATIN I
CC NUMBER OF SEQUENCES: 11
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: HUMAN GENOME SCIENCES, INC.
CC STREET: 9410 KEY WEST AVENUE
CC CITY: ROCKVILLE
CC STATE: MD
CC COUNTRY: USA
CC ZIP: 20850
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/409.731A
CC FILING DATE: 24-MAR-1995
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: BENSON, ROBERT H
CC REGISTRATION NUMBER: 30,446
CC REFERENCE/DOCKET NUMBER: PF175
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 301-309-8504
CC TELEFAX: 301-309-8512
CC INFORMATION FOR SEQ ID NO: 11:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 132 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
SQ SEQUENCE 132 AA; 14909 MW; 94859 CN;

Query Match 51.9%; Score 54; DB 1; Length 132;
Best Local Similarity 62.5%; Pred. No. 7.55e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 4 KFLGTWKL 11
QY 1 KFLAVWKI 8
||||: ||:

RESULT 11
ID US-08-241-664B-14 STANDARD; PRT; 25 AA.
XX
AC xxxxxx
XX
DT
XX
XX
DE Sequence 14, Application US/08241664B
XX
XX Sequence 14, Application US/08241664B
CC Patent No. 5871909
CC GENERAL INFORMATION:
CC APPLICANT: VOORHEES, JOHN J.
CC APPLICANT: ASTROM, ANDERS
CC APPLICANT: PATTERSON, ULRIKA
CC APPLICANT: TAVAKKOL, AMIR
CC TITLE OF INVENTION: HUMAN CRABP-I AND CRABP-II
CC NUMBER OF SEQUENCES: 18
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: HARNESS, DICKEY & PIERCE, P.L.C.
CC STREET: PO Box 828
CC CITY: Bloomfield Hills
CC STATE: Michigan
CC COUNTRY: United States of America
CC ZIP: 48303
CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/241,664B
CC FILING DATE: May 11, 1994
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Smith, Deann F.
CC REGISTRATION NUMBER: 36,683
CC REFERENCE/DOCKET NUMBER: 2115-00676COD
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (810) 641-1600
CC TELEX: (810) 641-0270
CC TELEFAX: 287637
CC INFORMATION FOR SEQ ID NO: 14:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 25 amino acids
CC TYPE: amino acid
CC TOPOLOGY: unknown
CC MOLECULE TYPE: peptide
CC SEQUENCE 25 AA; 2950 MW; 3278 CN;
CC
CC Query Match 51.0%; Score 53; DB 2; Length 25;
CC Best Local Similarity 41.7%; Pred. No. 9.40e+01;
CC Matches 5; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
CC
Db 2 NFEWNKIIIXSE 13
QY 1 KFLAVWKITYKD 12
RESULT 12
ID US-08-468-709B-14 STANDARD; PRT; 25 AA.
XX XXXXXX
CC
CC Sequence 14, Application US/08468709B
CC Patent No. 5654137
CC GENERAL INFORMATION:
CC APPLICANT: Astrom, Anders
CC APPLICANT: Voorhees, John
CC APPLICANT: Patterson, Ulrika
CC APPLICANT: Tavakkol, Amir
CC TITLE OF INVENTION: HUMAN CRABP-I AND CRABP-II
CC NUMBER OF SEQUENCES: 18
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Harness, Dickey & Pierce, P.I.C.
CC STREET: PO Box 828
CC CITY: Bloomfield Hills
CC STATE: Michigan
CC COUNTRY: United States of America
CC ZIP: 48303
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/468,709B
CC FILING DATE: 05/06/95
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Smith, Deann F.
CC REGISTRATION NUMBER: 36,683
CC REFERENCE/DOCKET NUMBER: 2115-00676DVF
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (810) 641-1600

CC TELEFAX: (810) 641-0270
CC TELEX: 287637
CC INFORMATION FOR SEQ ID NO: 14:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 25 amino acids
CC TYPE: amino acid
CC TOPOLOGY: unknown
CC MOLECULE TYPE: peptide
CC SEQUENCE 25 AA; 2950 MW; 3278 CN;
CC
CC Query Match 51.0%; Score 53; DB 1; Length 25;
CC Best Local Similarity 41.7%; Pred. No. 9.40e+01;
CC Matches 5; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
CC
Db 2 NFEWNKIIIXSE 13
QY 1 KFLAVWKITYKD 12
RESULT 13
ID US-07-879-617A-9 STANDARD; PRT; 635 AA.
XX XXXXXX
CC
CC Sequence 9, Application US/07879617A
CC Patent No. 5580775
CC GENERAL INFORMATION:
CC APPLICANT: Freneau Jr., Robert T.
CC APPLICANT: Caron, Marc G.
CC APPLICANT: Blakely, Randy D.
CC TITLE OF INVENTION: A High Affinity L-Proline Transporter
CC NUMBER OF SEQUENCES: 13
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Kilpatrick & Cody
CC STREET: 1100 Peachtree Street, Suite 2800
CC CITY: Atlanta
CC STATE: Georgia
CC COUNTRY: U.S.
CC ZIP: 30309
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/07/879,617A
CC FILING DATE: 19920501
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Pabst, Patrea L.
CC REGISTRATION NUMBER: 31,284
CC REFERENCE/DOCKET NUMBER: EMU109
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 404-815-6508
CC TELEFAX: 404-815-6555
CC INFORMATION FOR SEQ ID NO: 9:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 635 amino acids
CC TYPE: AMINO ACID
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC HYPOTHEetical: NO
CC ANTI-SENSE: NO
CC FRAGMENT TYPE: N-terminal
CC ORIGINAL SOURCE:
CC ORGANISM: Rattus
CC TISSUE TYPE: Brain - Proline Transporter
CC FEATURE:

CC NAME/KEY: Domain
 CC LOCATION: 46..65
 CC OTHER INFORMATION: /note= "Proposed transmembrane
 CC OTHER INFORMATION: domain."
 CC FEATURE:
 CC NAME/KEY: Domain
 CC LOCATION: 72..97
 CC OTHER INFORMATION: /note= "Proposed transmembrane
 CC OTHER INFORMATION: domain."
 CC FEATURE:
 CC NAME/KEY: Domain
 CC LOCATION: 117..137
 CC OTHER INFORMATION: /note= "Proposed transmembrane
 CC OTHER INFORMATION: domain."
 CC FEATURE:
 CC NAME/KEY: Domain
 CC LOCATION: 217..236
 CC OTHER INFORMATION: /note= "Proposed transmembrane
 CC OTHER INFORMATION: domain."
 CC FEATURE:
 CC NAME/KEY: Domain
 CC LOCATION: 243..264
 CC OTHER INFORMATION: /note= "Proposed transmembrane
 CC OTHER INFORMATION: domain."
 CC FEATURE:
 CC NAME/KEY: Domain
 CC LOCATION: 291..311
 CC OTHER INFORMATION: /note= "Proposed transmembrane
 CC OTHER INFORMATION: domain."
 CC FEATURE:
 CC NAME/KEY: Domain
 CC LOCATION: 322..345
 CC OTHER INFORMATION: /note= "Proposed transmembrane
 CC OTHER INFORMATION: domain."
 CC FEATURE:
 CC NAME/KEY: Domain
 CC LOCATION: 373..397
 CC OTHER INFORMATION: /note= "Proposed transmembrane
 CC OTHER INFORMATION: domain."
 CC FEATURE:
 CC NAME/KEY: Domain
 CC LOCATION: 424..443
 CC OTHER INFORMATION: /note= "Proposed transmembrane
 CC OTHER INFORMATION: domain."
 CC FEATURE:
 CC NAME/KEY: Domain
 CC LOCATION: 456..479
 CC OTHER INFORMATION: /note= "Proposed transmembrane
 CC OTHER INFORMATION: domain."
 CC FEATURE:
 CC NAME/KEY: Domain
 CC LOCATION: 500..519
 CC OTHER INFORMATION: /note= "Proposed transmembrane
 CC OTHER INFORMATION: domain."
 CC FEATURE:
 CC NAME/KEY: Domain
 CC LOCATION: 536..559
 CC OTHER INFORMATION: /note= "Proposed transmembrane
 CC OTHER INFORMATION: domain."
 CC FEATURE:
 CC NAME/KEY: Active-site
 CC LOCATION: 76..77
 CC OTHER INFORMATION: /note= "Leucine zipper motif"
 CC FEATURE:
 CC NAME/KEY: Active-site
 CC LOCATION: 83..84
 CC OTHER INFORMATION: /note= "Leucine zipper motif"
 CC FEATURE:
 CC NAME/KEY: Active-site
 CC LOCATION: 90..91
 CC OTHER INFORMATION: /note= "Leucine zipper motif"
 CC FEATURE:
 CC NAME/KEY: Active-site

CC LOCATION: 97..98
 CC OTHER INFORMATION: /note= "Leucine zipper motif"
 CC SEQUENCE 635 AA: 70857 MW; 2226371 CN;
 SQ
 Query Match 51.0%; Score 53; DB 1; Length 635;
 Best Local Similarity 85.7%; Pred. NO. 9.40e+01;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Db 106 LAVWKIS 112
 QY 3 LAVWKIT 9
 RESULT 14
 ID US-08-753-985-9 STANDARD; PRT; 635 AA.
 XX
 AC xxxxxx
 XX
 DT
 XX
 XX
 DE Sequence 9, Application US/08753985
 XX Sequence 9, Application US/08753985
 CC Patent No. 5759788
 CC GENERAL INFORMATION:
 CC APPLICANT: Fremont Jr., Robert T.
 CC APPLICANT: Caron, Marc G.
 CC APPLICANT: Blakely, Randy D.
 CC TITLE OF INVENTION: A High Affinity L-Proline Transporter
 CC NUMBER OF SEQUENCES: 13
 CC CORRESPONDENCE ADDRESS:
 CC ADDRESSEE: Kilpatrick & Cody
 CC STREET: 1100 Peachtree Street, Suite 2800
 CC CITY: Atlanta
 CC STATE: Georgia
 CC COUNTRY: U.S.
 CC ZIP: 30309
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Floppy disk
 CC COMPUTER: IBM PC compatible
 CC OPERATING SYSTEM: PC-DOS/MS-DOS
 CC SOFTWARE: Patent Release #1.0, Version #1.25
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: US/08/753,985
 CC FILING DATE: 03-DEC-1996
 CC CLASSIFICATION: 536
 CC PRIOR APPLICATION DATA:
 CC APPLICATION NUMBER: US 07/879617
 CC FILING DATE: 01-MAY-1992
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: Pabst, Patrea L.
 CC REGISTRATION NUMBER: 31,284
 CC REFERENCE/DOCKET NUMBER: EMU109
 CC TELECOMMUNICATION INFORMATION:
 CC TELEPHONE: 404-815-6508
 CC TELEFAX: 404-815-6555
 CC INFORMATION FOR SEQ ID NO: 9:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 635 amino acids
 CC TYPE: amino acid
 CC STRANDEDNESS: single
 CC TOPOLOGY: linear
 CC MOLECULE TYPE: protein
 CC HYPOTHETICAL: NO
 CC ANTI-SENSE: NO
 CC FRAGMENT TYPE: N-terminal
 CC ORIGINAL SOURCE:
 CC ORGANISM: Rattus
 CC TISSUE TYPE: Brain - Proline Transporter
 CC FEATURE:
 CC NAME/KEY: Domain
 CC LOCATION: 46..65
 CC OTHER INFORMATION: /note= "Proposed transmembrane

CC OTHER INFORMATION: domain."
CC FEATURE:
CC NAME/KEY: Domain
CC LOCATION: 72..97
CC OTHER INFORMATION: /note= "Proposed transmembrane
CC OTHER INFORMATION: domain."
CC FEATURE:
CC NAME/KEY: Domain
CC LOCATION: 117..137
CC OTHER INFORMATION: /note= "Proposed transmembrane
CC OTHER INFORMATION: domain."
CC FEATURE:
CC NAME/KEY: Domain
CC LOCATION: 217..236
CC OTHER INFORMATION: /note= "Proposed transmembrane
CC OTHER INFORMATION: domain."
CC FEATURE:
CC NAME/KEY: Domain
CC LOCATION: 243..264
CC OTHER INFORMATION: /note= "Proposed transmembrane
CC OTHER INFORMATION: domain."
CC FEATURE:
CC NAME/KEY: Domain
CC LOCATION: 291..311
CC OTHER INFORMATION: /note= "Proposed transmembrane
CC OTHER INFORMATION: domain."
CC FEATURE:
CC NAME/KEY: Domain
CC LOCATION: 322..345
CC OTHER INFORMATION: /note= "Proposed transmembrane
CC OTHER INFORMATION: domain."
CC FEATURE:
CC NAME/KEY: Domain
CC LOCATION: 373..397
CC OTHER INFORMATION: /note= "Proposed transmembrane
CC OTHER INFORMATION: domain."
CC FEATURE:
CC NAME/KEY: Domain
CC LOCATION: 424..443
CC OTHER INFORMATION: /note= "Proposed transmembrane
CC OTHER INFORMATION: domain."
CC FEATURE:
CC NAME/KEY: Domain
CC LOCATION: 456..479
CC OTHER INFORMATION: /note= "Proposed transmembrane
CC OTHER INFORMATION: domain."
CC FEATURE:
CC NAME/KEY: Domain
CC LOCATION: 500..519
CC OTHER INFORMATION: /note= "Proposed transmembrane
CC OTHER INFORMATION: domain."
CC FEATURE:
CC NAME/KEY: Domain
CC LOCATION: 536..559
CC OTHER INFORMATION: /note= "Proposed transmembrane
CC OTHER INFORMATION: domain."
CC FEATURE:
CC NAME/KEY: Active-site
CC LOCATION: 76..77
CC OTHER INFORMATION: /note= "Leucine zipper motif"
CC FEATURE:
CC NAME/KEY: Active-site
CC LOCATION: 83..84
CC OTHER INFORMATION: /note= "Leucine zipper motif"
CC FEATURE:
CC NAME/KEY: Active-site
CC LOCATION: 90..91
CC OTHER INFORMATION: /note= "Leucine zipper motif"
CC FEATURE:
CC NAME/KEY: Active-site
CC LOCATION: 97..98
CC OTHER INFORMATION: /note= "Leucine zipper motif"
CC SEQUENCE 635 AA; 70857 MW; 2226371 CN;

Query Match 51.0%; Score 53; DB 2; Length 635;
Best Local Similarity 85.7%; Pred. No. 9,40e+01;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 106 LAVWKIS 112
QY 3 LAVWKIT 9
RESULT 15
ID US-08-753-985-8 STANDARD; PRT; 667 AA.
XX
AC xxxxxx
XX
DT
XX
DE
XX
CC Sequence 8, Application US/08753985
CC Sequence 8, Application US/08753985
CC Patent No. 5759788
CC GENERAL INFORMATION:
CC APPLICANT: Freneau Jr., Robert T.
CC APPLICANT: Caron, Marc G.
CC APPLICANT: Blakely, Randy D.
CC TITLE OF INVENTION: A High Affinity L-Proline Transporter
CC NUMBER OF SEQUENCES: 13
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Kilpatrick & Cody
CC STREET: 1100 Peachtree Street, Suite 2800
CC CITY: Atlanta
CC STATE: Georgia
CC COUNTRY: U.S.
CC ZIP: 30309
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/753,985
CC FILING DATE: 03-DEC-1996
CC CLASSIFICATION: 536
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/879617
CC FILING DATE: 01-MAY-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Pabst, Patrea L.
CC REGISTRATION NUMBER: 31,284
CC REFERENCE/DOCKET NUMBER: EMU109
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 404-815-6508
CC TELEFAX: 404-815-6555
CC INFORMATION FOR SEQ ID NO: 8:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 667 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC HYPOTHETICAL: NO
CC ANTI-SENSE: NO
CC FRAGMENT TYPE: N-terminal
CC ORIGINAL SOURCE:
CC ORGANISM: Rattus
CC TISSUE TYPE: Brain
CC IMMEDIATE SOURCE:
CC LIBRARY: rat forebrain cDNA library
CC CLONE: rTB2-2-20
CC FEATURE:
CC NAME/KEY: Domain
CC LOCATION: 76..95
CC OTHER INFORMATION: /note= "Membrane-spanning domain"

CC FEATURE:
CC NAME/KEY: Domain
CC LOCATION: 103..127
CC OTHER INFORMATION: /note= "Membrane-spanning domain"
CC FEATURE:
CC NAME/KEY: Domain
CC LOCATION: 147..167
CC OTHER INFORMATION: /note= "Membrane-spanning domain"
CC FEATURE:
CC NAME/KEY: Domain
CC LOCATION: 247..266
CC OTHER INFORMATION: /note= "Membrane-spanning domain"
CC FEATURE:
CC NAME/KEY: Domain
CC LOCATION: 272..294
CC OTHER INFORMATION: /note= "Membrane-spanning domain"
CC FEATURE:
CC NAME/KEY: Domain
CC LOCATION: 321..341
CC OTHER INFORMATION: /note= "Membrane-spanning domain"
CC FEATURE:
CC NAME/KEY: Domain
CC LOCATION: 353..375
CC OTHER INFORMATION: /note= "Membrane-spanning domain"
CC FEATURE:
CC NAME/KEY: Domain
CC LOCATION: 404..427
CC OTHER INFORMATION: /note= "Membrane-spanning domain"
CC FEATURE:
CC NAME/KEY: Domain
CC LOCATION: 454..473
CC OTHER INFORMATION: /note= "Membrane-spanning domain"
CC FEATURE:
CC NAME/KEY: Domain
CC LOCATION: 487..509
CC OTHER INFORMATION: /note= "Membrane-spanning domain"
CC FEATURE:
CC NAME/KEY: Domain
CC LOCATION: 530..549
CC OTHER INFORMATION: /note= "Membrane-spanning domain"
CC FEATURE:
CC NAME/KEY: Domain
CC LOCATION: 567..589
CC OTHER INFORMATION: /note= "Membrane-spanning domain"
CC FEATURE:
CC NAME/KEY: Region
CC LOCATION: 44..45
CC OTHER INFORMATION: /note= "protein kinase C phosphorylation site"
CC FEATURE:
CC NAME/KEY: Region
CC LOCATION: 71..72
CC OTHER INFORMATION: /note= "cAMP-dependent protein kinase phosphorylation site"
CC FEATURE:
CC NAME/KEY: Region
CC LOCATION: 269..270
CC OTHER INFORMATION: /note= "protein kinase C phosphorylation site"
CC FEATURE:
CC NAME/KEY: Region
CC LOCATION: 630..631
CC OTHER INFORMATION: /note= "protein kinase C phosphorylation site"
CC FEATURE:
CC NAME/KEY: Region
CC LOCATION: 106..127
CC OTHER INFORMATION: /note= "Leucine zipper motif"
CC FEATURE:
CC NAME/KEY: Region
CC LOCATION: 212..213
CC OTHER INFORMATION: /note= "N-linked glycosylation site"
CC

SQ SEQUENCE 667 AA; 74584 MW; 2458833 CN;

Query Match 51.0%; Score 53; DB 2: Length 667;
Best Local Similarity 85.7%; Pred. NO: 9.40e+01;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 136 LAVWKIS 142
|||||:
QY 3 LAVWKII 9

Search completed: Tue Aug 17 16:04:43 1999
Job time : 7 secs.

This Page blank (uspto)